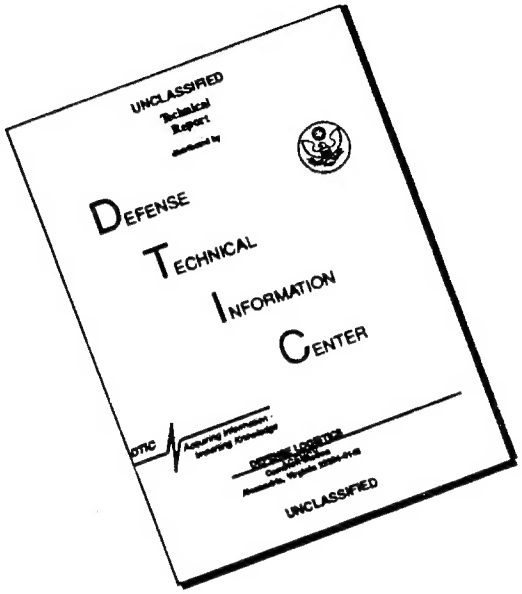
AFOSR-TR- 95 - 0252

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: University of Pennsylvania; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: AFOSR/NL
6c. ADDRESS (City, State and ZIP Code): Office of Research Administration, Philadelphia, PA 19104	7b. ADDRESS (City, State and ZIP Code): 110 Duncan Ave, Suite B115, Bolling AFB DC 20332-0001
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR - Life Science Directorate; 8b. OFFICE SYMBOL (If applicable): NL	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR 91-0082
8c. ADDRESS (City, State and ZIP Code): Bolling AFB, DC 20332	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO
61102F	2313	A5	

11. TITLE (Include Security Classification): Multidimensional Signal Coding in the Visual System

12. PERSONAL AUTHOR(S): Buchsbaum, Gershon

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Final	FROM 10/90 TO 10/94	95/3/17	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

See atch.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☑ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Dr John F. Tangney	202-767-5021	NL

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

SECURITY CLASSIFICATION OF THIS PAGE

AFOSR 91-0082

MULTIDIMENSIONAL SIGNAL CODING IN THE VISUAL SYSTEM

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements either expressed or implied, of the Air Force Office of Scientific Research or the US Government

AFOSR 91-0082

MULTIDIMENSIONAL SIGNAL CODING IN THE VISUAL SYSTEM

The purpose of the research was to investigate the significance of natural images in understanding early image coding in the visual system. The main objectives were:

1. To study the image processing capabilities in early vision and how they modify the space, time and color properties of an image and the efficiency of these image processing operations.

2. To identify key attributes of natural image signals which are sensitive to modification and filtering in early vision and how these attributes are transformed for encoding beyond early vision.

The research addressed a number of issues in early vision including the relationship between color, spatial and temporal properties of images and concentrated in three main thrusts.

1. The match between neural pathways in early vision and their underlying retinal architecture and spatial and temporal properties of images was investigated. Visual tracking was incorporated into the study as contributing important modifications to the incoming spectrum of the image. The results indicate that the combination of visual tracking together with specialized neural pathways in the retina makes the coding of the spatial and temporal features of images efficient. Bio-encoding of images in the early visual system is optimal in the sense that the retinal neural pathways are tuned to intrinsic properties of natural image sequences.

2. The multilayered retinal architecture and how it mediates signal propagation and prevents distortions as the signal propagates through retinal cell layers was investigated. The research incorporated anatomical and functional details of retinal architecture including different cell densities in different retinal cell layers, cell-to-cell variations, and how cell arrays sample and propagate the image. These properties were incorporated into a multi-stage signal processing model. Understanding of the hierarchical multi-layerd signal processing strategy of the retina revealed the role of various components of retinal anatomical architecture. Together with the optics of the eye, retinal architecture provides a means to preserve the image and prevent distortions in it.

3. Color constancy, or the ability of the visual system to perceive color independently of the ambient illumination, was investigated in the context of a biologically-based neural network. In particular, the role of retinal adaptation and higher level visual operations in mediating color constancy was investigated. The study incorporated properties of individual cells and how they combine to make complex color and spatial operations. The neural network simulations indicate how early visual stages complement each other to compensate and maintain relatively constant color perception under conditions of varying illumination and spatial context in the image.

The research was reported in papers published in scientific journals which are included as appendices A-H. The following pages include a list of the publications as well as other activities, and the personnel involved.

PUBLICATIONS

PAPERS IN SCIENTIFIC JOURNALS:

Eckert, Michael, P., Buchsbaum, Gershon, Watson, Andrew, B., Separability of Spatiotemporal Spectra of Image Sequences, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, pp. 1210-1213, 1992.

Levitan, Bennett, S., Buchsbaum, Gershon, Conversion Between Parallel and Hierarchic Architecture Analysis Multirate Filter Banks, IEEE Transactions on Signal Processing, Vol. 40, pp. 2837-2841, 1992

Levitan, Bennett, S., Buchsbaum, Gershon, Complexity and Filter Memory Requirements in Scaled Gaussian Hierarchic and Parallel Filter Banks, J. Visual Communications and Image Representation, Vol. 4, pp. 187-195 1993

Eckert, Michael, P., Buchsbaum, Gershon, Efficient Coding of Natural Time Varying Images in the Early Visual System, Philosophical Transactions of the Royal Society (London), Series. B, Vol. 339, pp. 385-395 1993

Levitan, Bennett, S., Buchsbaum, Gershon, Signal Sampling and Propagation through Multiple Cell Layers in the Retina: Modeling and Analysis Using Multirate Filtering, Journal of the Optical Society of America, Series A, Vol. 7, pp. 1463-1480, 1993

Eckert, Michael, P., Buchsbaum, Gershon, Effect of Tracking Strategies on the Velocity Structure of Image Sequences, Journal of the Optical Society of America, Series A, Vol. 7, pp. 1582-1584, 1993

Courtney, Susan, M., Finkel, Leif, H., Buchsbaum, Gershon, A Multi-Stage neural network for Color Constancy and Color Induction, IEEE Transactions on Neural Networks, in press, 1995

Courtney, Susan, M., Finkel, Leif, H., Buchsbaum, Gershon, Network Simulations of Retinal and Cortical Contributions to Color Constancy, Vision Research Vol. 35, pp. 413-434, 1995

BOOK CHAPTERS:

Buchsbaum, Gershon, Visual Systems Considerations in the Coding of Natural Color Images, in Digital Images and Human Vision: A. B. Watson, editor, MIT Press, 99-108, 1993.

Eckert, Michael, P., Buchsbaum, Gershon, The Significance of Eye Movement and Image Acceleration for Coding Television Images, in Digital Images and Human Vision: A. B. Watson, editor, MIT Press, 89-98, 1993.

CONFERENCES, PROCEEDINGS, SYMPOSIA, WORKSHOPS:

Buchsbaum, Gershon, The Relationship Between Space and Color in Natural Images, National Research Council Committee on Vision, Workshop on Visual Factors in Electronic Image Communication, Woods Hole, MA, (1991).

Levitan, Bennett, S., Buchsbaum, Gershon, Architecture-Dependent Properties of Analysis Multirate Filter Banks, Proceedings of . International Conference on Acoustics Speech and Signal Processing IEEE, ICASSP Vol. 3, pp. 1805-1808 (1991).

Eckert, Michael, P., Buchsbaum, Gershon, M and P Cells Are Matched to the Spatiotemporal Image Spectrum Modified by Eye Movements, Investigative Ophthalmology & Visual Science (ARVO), Vol. 32 pp. 841, 1991.

Derrico, Joel, B., Buchsbaum, Gershon, Efficient Image Coding Operations in Space and Color and their Correlates in the Early Visual System, Investigative Ophthalmology & Visual Science (ARVO), Vol. 32 pp. 1267, 1991.

Eckert, Michael, P., Buchsbaum, Gershon, The Effect of Eye-Movement Tracking Strategies on the Velocity Distribution of the Retinal Image, Annual Meeting of the Optical Society of America Technical Digest Series Vol. 23, pp. 119 (1992).

Courtney, Susan, M., Buchsbaum, Gershon, Finkel, Leif, H., Cone Adaptation and Cortical Silent Surrounds Cooperate to Produce Color Constancy and Color Induction, Annual Meeting of the Optical Society of America Technical Digest Series Vol. 23, pp. 63 (1992).

Buchsbaum, Gershon, The Basic Building Blocks of Color Vision: A Generalized View of the Opponent Colors Transformation, Advances in Color Vision, Optical Society of America, Vol. 4 pp. 84-86 (1992).

"Biologically-Based Neural Network Model of Color Constancy and Color Contrast," S. M. Courtney, G. Buchsbaum and L. H. Finkel, IEEE International Joint Conference on Neural Networks, Vol. 4, pp. 55-60 (1992).

Eckert, Michael, P., Buchsbaum, Gershon, The Relationship Between Retinal Receptor Packing and Tracking Eye Movement, Investigative Ophthalmology & Visual Science (ARVO) Vol. 33 pp. 1144, 1992.

Courtney, Susan, M., Buchsbaum, Gershon, Finkel, Leif, H., Color Constancy and Color Contrast in a Physiologically-Based Network Model, Investigative Ophthalmology & Visual Science (ARVO) Vol. 33 pp. 704, 1992.

Courtney, Susan, M., Buchsbaum, Gershon, Finkel, Leif, H., The Effects of Color-Opponent and Cone-Specific Processing Stages on Color and Brightness Perception, Investigative Ophthalmology & Visual Science (ARVO) Vol. 34, pp. 746 (1993)

Levitan, Bennett, S., Buchsbaum, Gershon, Multirate Filtering: A New Approach to Modeling Signal Sampling and Propagation in Multiple Retinal Cell Layers, Investigative Ophthalmology & Visual Science (ARVO), Vol. 34, pp. 783 (1993)

Courtney, Susan, M., Finkel, Leif, H., Buchsbaum, Gershon, The Effect of Opponent processing and Spatial Integration on 'Equivalent surrounds' Investigative Ophthalmology & Visual Science (ARVO), Vol. 35, pp. 1637, 1994.

PARTICIPATING PERSONNEL:

Faculty:

Buchsbaum, Gershon
Finkel, Leif, H.

Graduate students and Ph.D. thesis titles

Eckert, Michael, P., (Ph.D.) Processing and Coding of Natural Time-Varying Images in the Retina." (1992)

Courtney, Susan, M., (Ph.D.) Retinal and Cortical Contributions to Color Constancy and Color Induction in a Multi-Stage Network (1993).

Levitan, Bennett, S., (M.D./Ph.D.), Image Propagation through Multiple Retinal Cell Layers: Multirate Filter-Based Modeling and Analysis in the Cat Retina (1994)

APPENDICES:

The appendices are arranged in three groups corresponding to the research thrusts.

Group 1: Spatiotemporal visual image coding

Appendix A: Efficient Coding of Natural Time Varying Images in the Early Visual System, Philosophical Transactions of the Royal Society (London), Series. B, Vol. 339, pp. 385-395 1993

Appendix B: Effect of Tracking Strategies on the Velocity Structure of Image Sequences, Journal of the Optical Society of America, Series A, Vol. 7, pp. 1582-1584, 1993

Appendix C: Separability of Spatiotemporal Spectra of Image Sequences, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, pp. 1210-1213, 1992.

Group 2: Signal propagation in the retina

Appendix D: Signal Sampling and Propagation through Multiple Cell Layers in the Retina: Modeling and Analysis Using Multirate Filtering, Journal of the Optical Society of America, Series A, Vol. 7, pp. 1463-1480, 1993

Appendix E: Conversion Between Parallel and Hierarchic Architecture Analysis Multirate Filter Banks, IEEE Transactions on Signal Processing, Vol. 40, pp. 2837-2841, 1992

Appendix F: Complexity and Filter Memory Requirements in Scaled Gaussian Hierarchic and Parallel Filter Banks, J. Visual Communications and Image Representation, Vol. 4, pp. 187-195 1993

Group 3: Color constancy and interactions of space and color

Appendix G: Network Simulations of Retinal and Cortical Contributions to Color Constancy, Vision Research Vol. 35, pp. 413-434, 1995

Appendix H: A Multi-Stage neural network for Color Constancy and Color Induction, IEEE Transactions on Neural Networks, in press, 1995

Appendices

Group 1: Spatiotemporal visual image coding

Appendix A: Efficient Coding of Natural Time Varying Images in the Early Visual System, Philosophical Transactions of the Royal Society (London), Series. B, Vol. 339, pp. 385-395 1993

Appendix B: Effect of Tracking Strategies on the Velocity Structure of Image Sequences, Journal of the Optical Society of America, Series A, Vol. 7, pp. 1582-1584, 1993

Appendix C: Separability of Spatiotemporal Spectra of Image Sequences, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, pp. 1210-1213, 1992.

Efficient coding of natural time varying images in the early visual system

MICHAEL P. ECKERT AND GERSHON BUCHSBAUM†

Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, 220 S. 33rd Street, Philadelphia, Pennsylvania 19104-6315, U.S.A.

SUMMARY

We investigate the hypothesis that the early visual system efficiently codes natural time varying images, first by tracking part of the image, then by matching the spatiotemporal properties of the neural pathway to those of the tracked image. A representation for the time varying image is formulated which consists of two spatiotemporal components, a velocity field component and a stationary component. We show, using digitized sequences of natural images, that the spatiotemporal spectrum and other attributes of the image markedly differ before and after tracking. The temporal frequency bandwidth and velocity distribution of the velocity field component are diminished in the region of tracking and broaden with increasing eccentricity from this region. On the other hand, the spectrum of the stationary component is unaffected by tracking. Comparison of the properties of the tracked image to those of the M and P pathways suggests that each pathway transmits different attributes of the tracked image. A retinal architecture which varies with eccentricity also matches the properties of the tracked image.

1. INTRODUCTION

Natural images contain spatiotemporal information comprised of motion and other time varying details such as flicker. Motion in the retinal image includes object motion in the visual scene, observer motion, and eye motion. Image motion presents a significant problem for efficient coding and representation of images in the visual system. The visual system must code and interpret the visual scene while accounting for objects moving at velocities which may exceed the temporal limitations of visual system processing.

The visual system confronts this complex time varying signal with two spatiotemporal mechanisms: eye movements and the spatiotemporal filter arrays known as the M and P pathways. In this paper, we examine how eye movements and the M and P pathways conjoin to make an efficient coder of the time varying image. Eye movements limit the temporal bandwidth of images by reducing the range of velocities reaching the fovea. M and P pathways efficiently carry image components, modified by eye movements, for analysis at cortical levels. The spatiotemporal properties of ganglion and lateral geniculate cells which form the M and P pathways have been extensively investigated in recent years (Kaplan & Shapley 1982; Hicks *et al.* 1982; Derrington & Lennie 1984; Blakemore & Vital-Durand 1986; Crook *et al.* 1988; Lee *et al.* 1989*a*; Purpura *et al.* 1990), especially the role of these pathways in coding spatiotemporal image components (Shapley & Perry 1986; Merigan 1986; Merigan 1989; Merigan & Maunsell 1990; Merigan *et al.* 1991; and Schiller *et al.* 1990). Generally, the M pathway is associated with fast temporal changes and the P pathway with high spatial acuity and colour, although there is considerable overlap across spatial and temporal frequencies.

The idea that the visual system efficiently codes the visual scene is not a new one (Barlow 1961, 1981; Snyder *et al.* 1977; Srinivisan *et al.* 1982; Buchsbaum & Gottschalk 1983; Laughlin 1983; Field 1987; Tsukamoto *et al.* 1990; Watson 1990; Derrico & Buchsbaum 1991). Under the efficient coding hypothesis, the purpose of retinal processing is to transmit visual information as effectively as possible to higher visual centers. This means that the visual system optimizes its coding strategy, given the physiological constraints of limited dynamic range of nerves, noise, and limited spatial and temporal bandwidths.

A general block diagram of the coding system under investigation is presented in figure 1. The coder is comprised of two components, the pre-retinal eye movements, modelled as a linear time variant filter, and the retinal spatiotemporal pathways, modelled as linear time invariant filters. Investigation of efficiency and other properties of the coder requires an understanding of the signal environment in which it operates. For the visual system, the environment is an observer freely viewing natural images. We begin by investigating the spatiotemporal spectrum of the time varying image (the input, $I(\boldsymbol{u},t)$, in figure 1) and the effects of tracking on the spatiotemporal spectrum and other properties of the image (the signal at the retina, $I_r(\boldsymbol{u},t)$, in figure 1). A representation of natural images

† To whom correspondence should be sent.

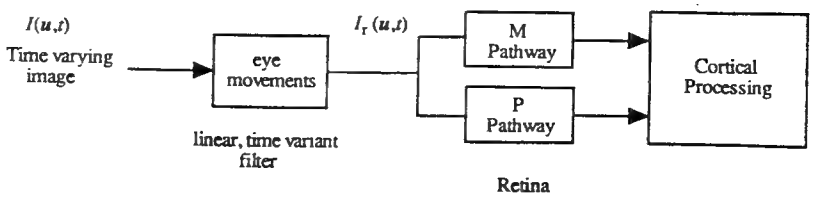


Figure 1. Block diagram of the flow of visual information through the visual system. The original image, $I(\boldsymbol{u},t)$, is filtered by eye movements. Eye movements are modeled as a linear time variant filter. The image reaching the retina, $I_r(\boldsymbol{u},t)$, is a filtered version of the original image. The M and P pathways, which are time invariant filters, further process the image before it reaches the cortex for analysis.

is formulated as a combination of a velocity field component and a stationary component which have markedly different spatiotemporal spectra. We then apply a tracking algorithm to natural image sequences to investigate how these two image components are affected. As expected, tracking reduces the temporal frequency bandwidth of the velocity field component at the point of tracking (Jain & Jain 1981; Girod 1987), but the temporal frequency bandwidth and velocity distribution broaden with increasing eccentricity from the point of tracking. The spectrum of the stationary component is not affected by tracking.

We discuss the properties of the M and P pathways and how they match the spatiotemporal components of images after tracking. We give special attention to calculating the effect of tracking in the region of tracking and at increasing eccentricities from it. This is needed to evaluate the advantage of a retinal architecture with a velocity tuning which changes with eccentricity.

2. SPATIOTEMPORAL SPECTRUM OF TIME VARYING IMAGES AND THE EFFECT OF TRACKING

(a) *Spectrum and velocity distribution of images*

Spatiotemporal variations in a visual scene generally arise from motion. On the image plane of the retina, motion can be approximated with a two-dimensional velocity field. The velocity field assigns a translational velocity vector to each point in space, and characterizes time variations resulting from geometric motion in the scene, including rotation, dilation, and affine deformations commonly found with perspective projections of three-dimensional motion. The velocity field serves as a basis for many computational models of human motion processing (Adelson & Bergen 1985; van Santen & Sperling 1985; Watson & Ahumada 1985; Heeger 1987). However, the velocity field cannot account for spatiotemporal changes of the image such as flicker and the photometric effects of motion (Pentland, 1991). To include all spatiotemporal changes, we model images as a combination of two uncorrelated components, a velocity field component and a stationary component.

$$I(\boldsymbol{u},t) = I\left(\boldsymbol{u} - \int_{t_0}^{t} \boldsymbol{v}(\boldsymbol{u},t')\,\mathrm{d}t', t_0\right) + s_s(\boldsymbol{u},t), \qquad (1)$$

where $I(\boldsymbol{u},t)$ is the intensity at spatial point $\boldsymbol{u}$ and time t, $I(\boldsymbol{u},t_0)$ specifies the initial image intensity, $\boldsymbol{v}(\boldsymbol{u},t)$ is the velocity field assigning a velocity vector, $\boldsymbol{v} = (v_x, v_y)$, to each point of space and time, and $s_s(\boldsymbol{u},t)$ is the stationary component.

The stationary component can be formed by removing local translational motion. Conceptually, this operation is analogous to filtering the image with a space–time variant filter which removes the velocity field component. The residual spatiotemporal intensity variations, the stationary component, will consist of flicker of the illuminant, the photometric effects of motion, and the occlusion and disocclusion at the edges of moving objects. These spatiotemporal effects are biologically relevant. For example, photometric motion provides depth and three-dimensional structure information about the image (Pentland 1991), and occlusion effects provide information about the location of object edges and relative depth. While these effects are caused by object motion, they cannot be removed by translational shifts of image intensity, and thus cannot be incorporated into the velocity field component.

The space and time variant spatiotemporal spectrum derived from the model of equation (1) is

$$S(\boldsymbol{u},t,\boldsymbol{k},f) = S(\boldsymbol{k})\delta(f - \boldsymbol{v}(\boldsymbol{u},t)\cdot\boldsymbol{k}) + S_s(\boldsymbol{k},f), \qquad (2)$$

where $\delta(\,)$ is the Dirac delta function, $\boldsymbol{k} = (k_x, k_y)$ is a two-dimensional spatial frequency vector, f is temporal frequency, $S(\boldsymbol{k})$ is the spatial power spectrum of $I(\boldsymbol{u},t_0)$, and $S_s(\boldsymbol{k},f)$ is the spatiotemporal spectrum of the stationary component. By definition, the velocity field and stationary component are uncorrelated, so the spectrum is the sum of the two components. In local spatiotemporal regions, the energy of the velocity field component exists on a plane in the three dimensions of frequency space (2 dimensions spatial, 1 dimension temporal), where the local velocity determines the orientation of the plane (Watson 1983; Watson & Ahamada 1985). A highly ordered structure does not exist for the stationary component which is distributed throughout spatiotemporal frequency space. Figure 2 illustrates the differences between the spectra of these components in a local spatiotemporal neighborhood.

The spatiotemporal spectrum (equation 2) can also be defined in terms of a velocity distribution, which is a probability distribution (histogram) of velocities. The velocity distribution is formed by sampling the velocity field through time at every spatial point in

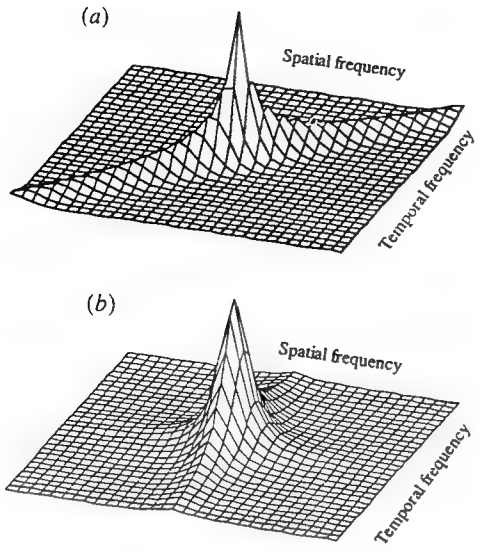


Figure 2. The velocity field and stationary components of time varying images occupy different regions of spatiotemporal frequency space. (*a*) The energy of translational motion lies along a line with a slope equal to velocity in spatiotemporal frequency space (a plane when spatial frequency is two dimensional). This figure illustrates an idealized case where the velocity field component has a constant translational velocity of 1 in the appropriate spatial and temporal frequency units. (*b*) The stationary component is modeled as the product of separable functions of spatial and temporal frequency which fall off inversely with spatial and temporal frequency.

the image. The result is a set of distributions, each representing the velocity variability in a region of space. Formation of the velocity distribution removes the functional dependence of time from the spatiotemporal spectrum, so it becomes a time averaged, but spatially localized, spatiotemporal spectrum. The formulation of the spectrum based on the velocity distribution is:

$$S(\boldsymbol{u},\boldsymbol{k},f) = h_v(\boldsymbol{u},\boldsymbol{v})S(\boldsymbol{k}) + S_s(\boldsymbol{k},f), \tag{3}$$

where $h_v(\boldsymbol{u},\boldsymbol{v})$ is the distribution of velocity, $\boldsymbol{v}$, at spatial point, $\boldsymbol{u}$, and $S(\boldsymbol{k})$ is the spatial power spectrum of the image.

(*b*) *Effect of tracking on the velocity field and spatiotemporal spectrum*

Eye movements can be modeled as a linear, time variant filter introduced between the image scene and the retina (figure 1). Inclusion of eye movements introduces a single time varying vector component to the velocity field. As eye movements are fully described by introducing a term to the velocity field, they do not modify the stationary component. The spatiotemporal spectrum after eye movements is:

$$S_e(\boldsymbol{u},t,\boldsymbol{k},f) = S(\boldsymbol{k})\delta(f - [\boldsymbol{v}(\boldsymbol{u},t) - \boldsymbol{v}_e(t)]\cdot\boldsymbol{k}) + S_s(\boldsymbol{k},f), \tag{4}$$

where $S_e(\boldsymbol{u},t,\boldsymbol{k},f)$ is the spatiotemporal spectrum which has been filtered by eye movements, and $\boldsymbol{v}_e(t)$ is the eye velocity at time, t.

Eye movements have the effect of shifting all velocities in the image by the eye velocity. In the case of tracking, eye velocity is set equal to the velocity at spatial point, $\boldsymbol{u}_0$.

$$\boldsymbol{v}_e(t) = \boldsymbol{v}(\boldsymbol{u}_0,t). \tag{5}$$

Tracking has the effect of minimizing the spread of the velocity distribution, $h_v(\boldsymbol{u},\boldsymbol{v})$, which decreases the temporal bandwidth of the signal in the neighborhood of the tracked point, $\boldsymbol{u}_0$. For perfect tracking, the velocity distribution becomes a delta function at point, $\boldsymbol{u}_0$, and we have the spatiotemporal spectrum,

$$S(\boldsymbol{u}_0,t,\boldsymbol{k},f) = S(\boldsymbol{k})\delta(f) + S_s(\boldsymbol{k},f). \tag{6}$$

Because tracking compensates for the velocity field component in the region of $\boldsymbol{u}_0$, the temporal variations are contributed by the stationary component, $S_s(\boldsymbol{k},f)$.

For points away from $\boldsymbol{u}_0$, the spectrum is weighted by the velocity distribution,

$$S(|\boldsymbol{u}-\boldsymbol{u}_0|,\boldsymbol{k},f) = h_v(|\boldsymbol{u}-\boldsymbol{u}_0|,\boldsymbol{v})S(\boldsymbol{k}) + S_s(\boldsymbol{k},f), \tag{7}$$

where $h_v(|\boldsymbol{u}-\boldsymbol{u}_0|,\boldsymbol{v})$ is a velocity distribution which broadens with increasing eccentricity, $|\boldsymbol{u}-\boldsymbol{u}_0|$. In the region around $\boldsymbol{u}_0$, tracking narrows the velocity distribution, thereby reducing the variability of the spatiotemporal spectrum and the temporal frequency bandwidth. With increasing eccentricity from the point of tracking, the spatiotemporal spectrum will have a broader velocity distribution and larger temporal frequency bandwidth. The degree to which tracking narrows the velocity distribution away from the point $\boldsymbol{u}_0$ depends on the spatial correlation of the velocity field, which is a measure of the change of the velocity field across space. For a highly correlated velocity field which changes slowly through space, tracking can reduce the velocity distribution at relatively large distances from the point of tracking.

Implementation of tracking invariably requires a feedback loop which estimates position and velocity from a time delayed input and past expectations (Stark *et al.* 1962; Lisberger *et al.* 1987; Steinman *et al.* 1990). This means that eye velocity can be set only to an estimated value of image velocity, rather than the true image velocity. As a result, the velocity distribution in the tracked region, $h_v(0,\boldsymbol{v})$, will have a spread related to the effectiveness of tracking. Tracking effectiveness is a signal related phenomenon, so that highly predictable motion will be more effectively tracked than unpredictable motion (Stark *et al.* 1962; Barnes & Lawson 1989). However, tracking of 'real world' motion can be quite accurate (Steinman *et al.* 1990), usually maintaining a foveal velocity of less than 1–2 deg s^{-1}.

3. THE EFFECT OF TRACKING ON DIGITIZED IMAGE SEQUENCES

We calculated the velocity field, velocity distribution, and spatiotemporal spectrum of four real world image sequences before and after tracking objects in the sequences. The sequences (256 pixels × 256 pixels × 64 frames at 8 bits per pixel, 30 frames per second with no scene cuts) were taken from a video disk which

Table 1. *Description of sequences*

sequence number	sequence description
IJ10833	Jungle scene with some three-dimensional object motion and a small amount of camera motion
IJ12426	Man walking. Some camera motion to keep man centred in visual field. The result is significant amounts of background motion
IJ01300	Man talking while moving head occasionally. No camera motion. Some slight motion in the background
IJ04454	Storm scene. No camera motion, but large amounts of non-rigid three-dimensional motion from waves. Intensity changes from lightning

contained scenes from movies. The frame rate of 30 frames per second limits the maximum estimated temporal frequency bandwidth of the images to 15 Hz. However, image energy drops off quickly with temporal frequency, and we found that signal energy is concentrated below 10 Hz. This suggests that any aliasing introduced by the sampling rate has little effect on the estimated spectrum. Each sequence was selected to contain varying levels of motion activity to form the broadest possible ensemble of images with the small sample size (see table 1). The velocity field for each frame was estimated by minimizing the squared difference between 24 pixel × 24 pixel blocks in two sequential frames of the sequence (Jain & Jain 1981). The same method was used to track selected regions in the sequence (see figure 3). While this (minimization of sum of squared diferences) algorithm is unlikely to be the method used by the visual system,

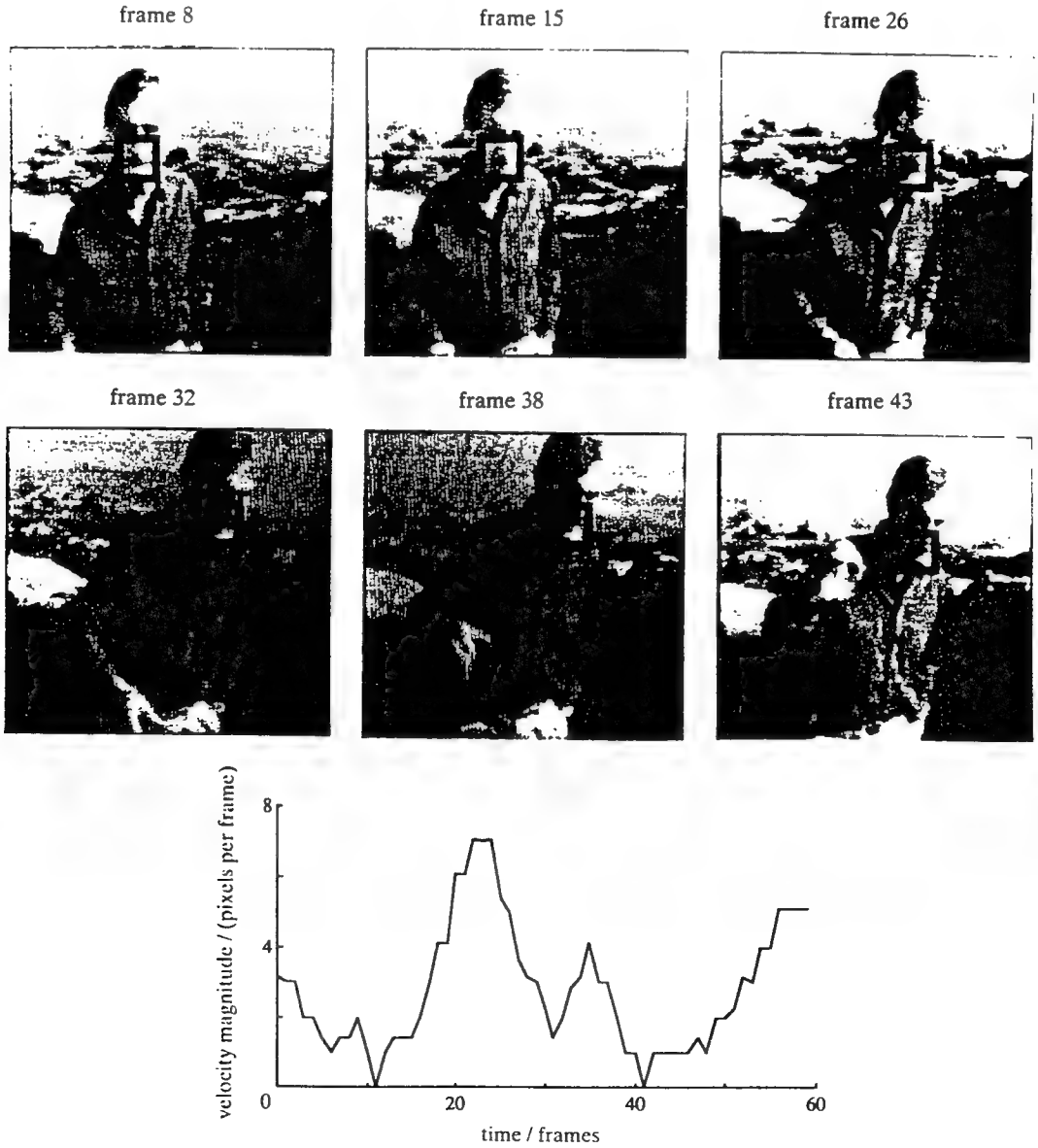


Figure 3. A typical image sequence over which the spatiotemporal statistics were analysed (sequence IJ12426). The black box indicates a region which was selected for tracking. The graph illustrates the magnitude of the velocity of the tracked region as a function of time.

it tracks a region of the image and keeps it centred, as does smooth pursuit.

We examined the effect of tracking on the velocity distribution in the region of tracking and at increasing eccentricities from that region. The velocity distribution of each 64 frame sequence was calculated from the velocity field as the frequency of occurrence of the velocity magnitude. To calculate the change of the velocity distribution with eccentricity, a region of interest was selected and tracked by shifting the entire image for each frame to maintain the tracked region in the same spatial location (figure 3). The velocity distribution and average velocity for the tracked image were then computed as a function of eccentricity from the point of tracking. The results are presented in figures 4 and 5. Before tracking, the velocity distribution and average velocity vary across the spatial extent of the image, but we found no trend from sequence to sequence. This is expected, as there is no reason why one part of the scene should experience more motion than any other part. When tracking is performed, the velocity distribution at the point of tracking is a delta function (for perfect tracking), and broadens with eccentricity. Higher velocities occur more frequently with increasing eccentricity from the region of tracking. The increase in average velocity with eccentricity was accompanied by a corresponding increase in the standard deviation, reflecting the fact that the variability of the velocity distribution increases with eccentricity. This result was consistent across all four sequences examined, although the degree to which tracking reduced the velocity distribution away from the point of tracking varied depending on the spatial correlation of the velocity field in that particular sequence. At the largest eccentricities, the velocity distribution after tracking can exceed the distribution before tracking. This can occur because the velocity distribution at large eccentricities is the vector sum of two uncorrelated velocity components, the image velocity and the velocity of eye movements.

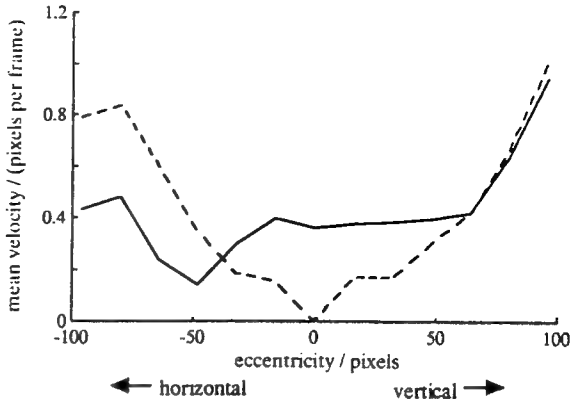


Figure 5. The average velocity in sequence IJ01300 before and after tracking (solid and dashed line respectively) as a function of eccentricity from the tracked point. The average was computed over the full 64 frames of the sequence. After tracking, the average velocity in the region of tracking drops to zero, and there is a regular increase in average velocity with eccentricity from the point of tracking. At large eccentricities, the average velocity after tracking can be larger than the average velocity before tracking.

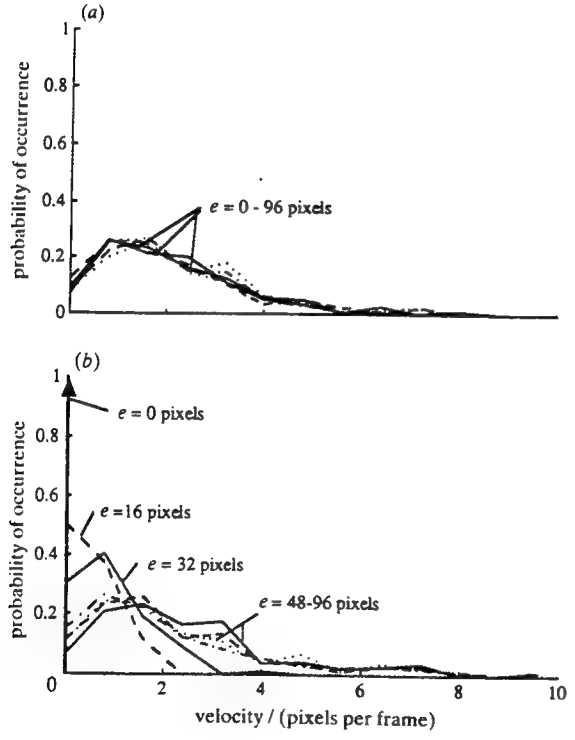


Figure 4. The velocity distribution of sequence IJ12426 as a function of eccentricity from the point of tracking. The distribution was computed from a 64 frame sequence. The curves represent the distribution at eccentricities of 0, 16, 32, 48, 64, 80, 96 pixels from the point of tracking. (*a*) Before tracking, the velocity distribution does not depend on eccentricity. The standard deviation for the curves is about 1.5 pixels per frame at all eccentricities. (*b*) After tracking, the distribution varies with eccentricity, from a delta function at the point of tracking, to a broad distribution at the largest eccentricity. The standard deviation increases with eccentricity and is 0, 0.51, 0.75, 1.6, 1.7, 1.71, 1.8 pixels per frame, respectively, for the eccentricities shown.

The changes in the velocity distribution as a result of tracking have corresponding effects in the spatiotemporal frequency spectrum (equation 7). In figure 6 we compare the spatiotemporal spectrum of images before and after tracking. The spectrum was computed in spatially and temporally localized blocks of size 32 pixels × 32 pixels × 16 frames at the point of tracking for a viewing distance of four screen heights (1 screen height = 256 pixels). For purposes of comparison, 1 pixel ≈ 1/15 deg and 1 pixel per frame ≈ 2 deg s^{-1} on a 256 pixel × 256 pixel image at a standard viewing distance of 4 screen heights with a frame rate of 30 frame per second. Most of the energy is concentrated below 10 Hz and 4 cycles per degree and diminishes quickly above these frequencies. After tracking, the temporal bandwidth of the spatiotemporal spectrum is greatly reduced. This is consistent with the changes in the velocity distribution after tracking (figure 4*b*). A reduction in the velocity distribution decreases the temporal bandwidth of the spatiotemporal spectrum.

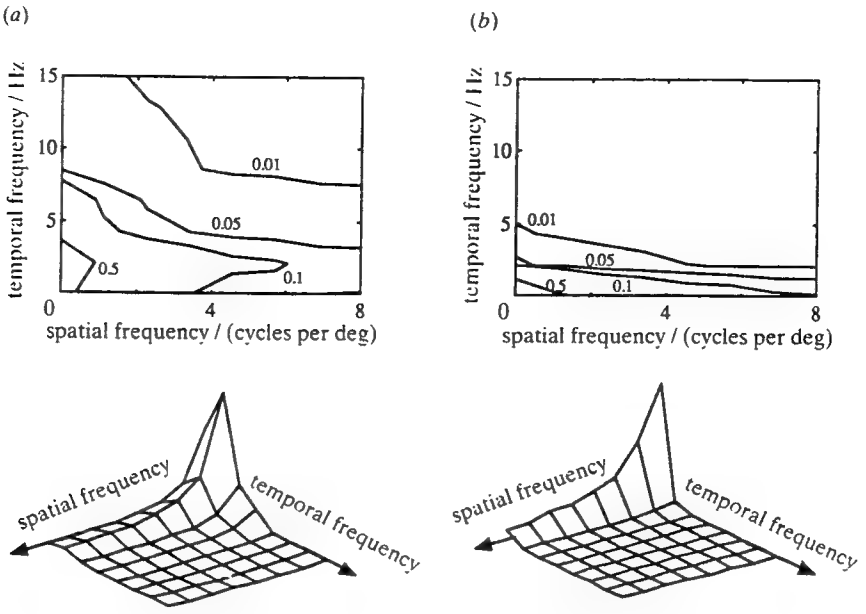


Figure 6. The spatiotemporal spectrum of the image before and after tracking computed for a $32 \times 32 \times 16$ spatiotemporal block from sequence IJ12426. The contour plots have lines at 0.01, 0.05, 0.1, 0.5 of the maximum values. The bottom figures are surface plots of the corresponding spectra. (*a*) Before tracking, the image has a large temporal bandwidth due to occasional large velocities. (*b*) After tracking, spatiotemporal energy is concentrated in lower temporal frequencies. The units of cycles per degree for spatial frequency axis were determined by using a viewing distance of 4 scene heights from the image. The range of spatial frequencies will vary for different viewing distances, but the range of temporal frequencies will not vary with viewing distance.

Figure 7 shows how tracking modifies the instantaneous temporal frequency bandwidth in the region of tracking. The instantaneous temporal frequency bandwidth was computed from the frame to frame correlation, ρ_τ, using a correlation model of the form $\rho_\tau = e^{-\alpha|\tau|}$ (Jayant & Noll 1984; Eckert *et al.* 1992). For a frame rate of 30 frames per second, the temporal bandwidth can be computed as $\alpha = -30 \log(\rho_\tau)$. Before tracking, the signal has a highly variable temporal frequency bandwidth, due to variable object velocity. After tracking, the temporal bandwidth is small and less variable. The temporal bandwidth is not zero after tracking, however, since tracking does not remove all time variations, only those which result from translational motion at the tracked point. Figure 7 highlights the two ways tracking affects temporal frequency bandwidth: (i) tracking greatly reduces the temporal frequency bandwidth during high velocity motion, and (ii) before tracking, the temporal frequency bandwidth fluctuates widely, depending on the velocity, but after tracking, the temporal frequency bandwidth in the region of tracking has only small fluctuations.

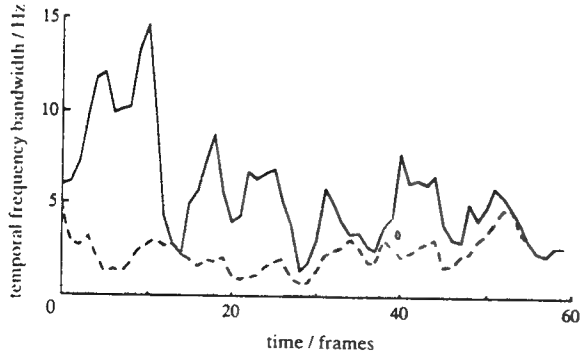


Figure 7. The effect of tracking on temporal frequency bandwidth for a 24 pixel $\times$ 24 pixel block from sequence IJ10833. The instantaneous temporal frequency bandwidth before tracking (solid line) and after tracking (dashed line) is computed from the frame to frame correlation. Before tracking, the signal has a highly variable temporal frequency bandwidth, due to variable object velocity. After tracking, the temporal bandwidth is small and does not vary significantly. Temporal frequency variations remain after tracking due to motion within the tracking region and the stationary component.

Tracking does not remove the velocity field component, but only shifts it to lower temporal frequencies. This shift will increase the energy share of the stationary component at high temporal frequencies. The relative share of the two components will vary from sequence to sequence and for different regions of tracking, depending on the amount of motion, and the degree to which time variations are represented by the velocity field component or by the stationary component. For the first three sequences in table 1, tracking of an object in the foreground reduces the average temporal bandwidth in the tracked region by 59%, 92%, and 56%, respectively. For these sequences, the large decrease of the temporal frequency bandwidth after tracking signifies that the velocity field component accounts for much of the signal energy at high temporal frequencies of the untracked image. The average temporal bandwidth of the last sequence (a

storm scene IJ04454) was reduced by only 22%. As tracking had little effect on the temporal bandwidth of this sequence, most time variations can be attributed to the stationary component. Large temporal intensity changes in this scene were due primarily to changing illuminant (lightning) and changing reflectance of light off ocean waves.

4. CODING BY THE VISUAL SYSTEM IN THE CONTEXT OF TRACKING

(*a*) *Advantages of tracking*

The velocity field of natural time varying images is signal dependent and variable. A scene may contain objects moving at high velocities, low velocities, or both. The corresponding spatiotemporal spectrum is also signal dependent and variable, with a large temporal bandwidth in the spatial regions which move at high velocities, and a small temporal bandwidth in slowly moving regions. A basic premise of coding theory is that a signal with a small bandwidth can be more efficiently coded than a signal with a large bandwidth (Jayant & Noll 1984). The coding efficiency is also affected by signal variability, because time-invariant coders (such as the retinal pathways) can be optimized only for a particular spectrum (Kassam & Poor 1983, 1985). The most efficiently coded signal is one with a small bandwidth and little or no variability. This corresponds to an image with little or no velocity, and thus a small temporal frequency bandwidth. Tracking with eye movements compensates for motion by matching eye velocity to the expected value of the image velocity in a region around the fovea. After tracking, the signal which actually reaches the retina (at the fovea) has a narrow velocity distribution and, therefore, a reduced temporal frequency bandwidth. A direct corollary of minimizing the temporal frequency bandwidth is a reduction in blur due to motion when the image is coded by fixed bandwidth time invariant channels. The role of eye movements in reducing blur was suggested before (Miller & Ludvigh 1962; Murphy 1978; Flipse *et al.* 1988) and follows from their role in the context of efficient coding.

Field (1987) showed that the spatial spectrum of natural images is scale invariant. This enables the visual system to use fixed, scene invariant, spatial filters to efficiently code a scene regardless of scale. The temporal spectrum and velocity distribution of natural time varying images are not scale invariant, and depend on the distance of moving objects from the observer. However, tracking maps the tracked region into the same temporal frequency and velocity distribution range regardless of the velocity (or scale) of the tracked region. Therefore, tracking provides a region of the retina with a virtually scale invariant signal in time and the coding advantages that accrue from the invariancy.

In addition to increased coding efficiency, tracking accentuates the importance of the stationary component in the temporal frequency domain in the region of tracking. Before tracking, this component is difficult to detect and isolate because it cannot easily be separated from the velocity field component. After tracking, the velocity field component is situated along the spatial frequency axis, and does not contribute to temporal variations, so the remaining temporal variations belong to the stationary term. Thus, perceptually important information associated with this component, such as flicker, photometric effects of motion, and motion edge effects, can be extracted more easily because of the removal of the velocity field component from high temporal frequencies. This argument does not hold in the periphery, however, because tracking only ensures reduction of the velocity field component in the region of tracking.

(*b*) *Retinal pathways and eccentricity dependent architecture are matched to the tracked image*

The second stage of the coder (figure 1) are the M and P pathways which operate on the tracked image, $I_r(\boldsymbol{u},t)$. These pathways and the underlying single cell units from which they are made have received considerable attention. Because of their significance in the present context, they are briefly reviewed here. The spatiotemporal filter properties of the M and P pathways are based on single cell properties of phasic and tonic cells from the retina and M and P cells from the LGN (Marrocco *et al.* 1982; Kaplan & Shapley 1982; Hicks *et al.* 1982; Derrington & Lennie 1984; Blakemore & Vital-Durand 1986; Crook *et al.* 1988; Lee *et al.* 1989b; Purpura *et al.* 1990). P (tonic) cells respond well to low temporal frequencies (below 5 Hz), whereas M (phasic) cells attenuate these frequencies. The spatial resolution of the P pathway is about three times higher than the M pathway at all eccentricities. This is due to receptive field center size and spatial sampling rates of the respective arrays (Merigan 1989; Merigan *et al.* 1991). The main characteristics of the pathways are summarized in table 2. The numbers in table 2 represent averages within the respective pathways rather than the response of any particular cell since there are large deviations among cells even in the same pathway (Hicks *et al.* 1982; Marrocco *et al.* 1982; Derrington & Lennie 1984).

Figure 8 illustrates the spatiotemporal transfer function of the M and P pathways inferred from the specifications in table 2. To obtain these responses, we fitted a frequency transfer function with the form of a spatial and temporal difference of Gaussians (Rohaly & Buchsbaum 1988; Rohaly 1988) and selected constants so as to meet the spatial and temporal frequency slopes and peaks in table 2.

$$\mathrm{RF}(|\boldsymbol{k}|,f,e) = [C_1 \mathrm{e}^{-\pi r_c(e)|\boldsymbol{k}|)^2} - S_1 \mathrm{e}^{-(\pi r_s(e)|\boldsymbol{k}|)^2}][C_2 \mathrm{e}^{-T_c f)^2} - S_2 \mathrm{e}^{-T_s f^2}], \quad (8)$$

where e is eccentricity, $r_c(e)$, $r_s(e)$ are the centre and surround sizes for receptive fields, and $\boldsymbol{k}$ and f are the spatial and temporal frequencies, respectively. T_c, T_s are temporal constants selected so as to provide peak temporal response at a specified frequency, and C_1, C_2, S_1, S_2 were selected so as to provide a specified response at low temporal and spatial frequencies. The

Table 2. *Spatial and temporal characteristics of the M and P pathways*

	M pathway	P pathway
spatial structure	centre-surround (relatively powerful surround)	centre surround (surrounds often have little power)
spatial resolution	one-third that of P cells (decreases with eccentricity)	three times that of M cells (decreases with eccentricity)
foveal spatial resolution	13 cycles per degree	40 cycles per degree
numbers of cells	10% of cells	80% of cells
spatial sampling rate	one-third that of P cells	three times that of M cells (80 samples per degree at fovea)
temporal frequency-peak	20 Hz: large variance between individual cells	10 Hz: large variance between individual cells
response at low frequencies	highly attenuated: phasic response to a step increase in light intensity	partially attenuated: tonic or sustained response to a step increase in light intensity
high frequency cutoff	up to 80 Hz	20–40 Hz
contrast sensitivity	high: eight times higher than P cells	low: eight times lower than M cells
speed (latency of response to visual stimulation)	fast: latency is about 24 ms at LGN for visual stimulation. Large variance	slow: latency is about 28 ms at LGN for visual stimulation. Large variance
suggested roles	carries information about quickly moving images with a low degree of spatial detail, such as flicker	carries slowly moving images with a high degree of spatial detail

centre and surround sizes for the receptive fields, $r_c(e)$ and $r_s(e)$, are assumed to increase in size with the inverse of the cortical magnification factor (Sakitt & Barlow 1982).

$$r_c(e) = r_c(0)(1+0.33e),\ r_s(e) = r_s(0)(1+0.33e). \quad (9)$$

The spatiotemporal frequency the velocity responses of M and P pathways can now be discussed in the context of the properties of images before and after tracking. Before tracking, the image spatiotemporal

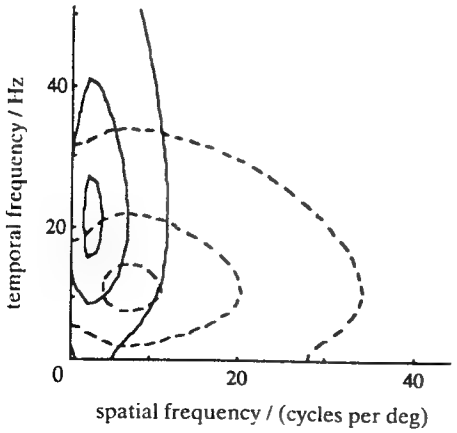


Figure 8. The spatiotemporal transfer function of the M and P pathways calculated from equation (8) with constants chosen to match details in table 2 (solid lines, M pathway; dashed lines, P pathway). Contours are at 0.1, 0.5, and 0.9 of the maximum value in the respective pathway. The P pathway is tuned to higher spatial frequencies and lower temporal frequencies than the M pathway, though there is considerable overlap.

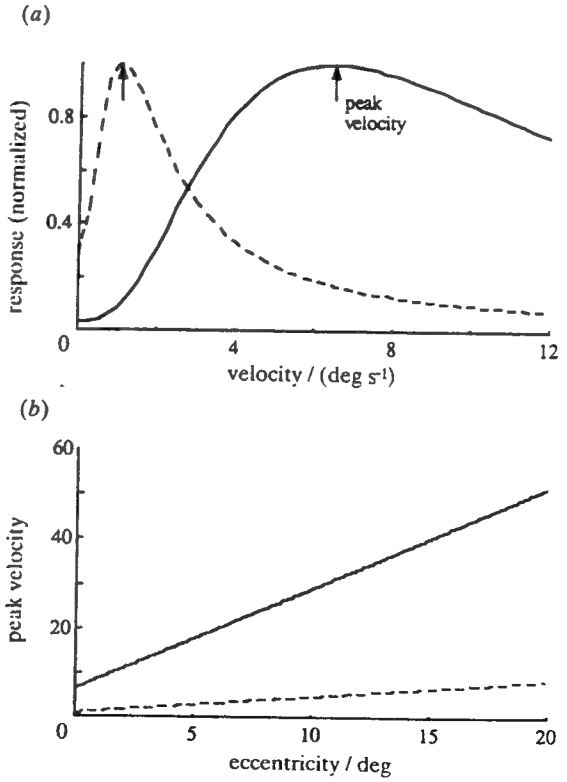


Figure 9. (*a*) The simulated response of M and P pathways at the fovea to a translating white noise stimulus as a function of velocity (solid line, M pathway; dashed line, P pathway). This is equivalent to integrating the frequency response over lines of constant velocity. The normalized peak response of the M and P pathways (arrows) is about 7 deg s^{-1} and 1 deg s^{-1}, respectively. (*b*) Peak velocity of the two pathways as a function of eccentricity. The peak velocity of the M pathway increases with eccentricity at a greater rate than the P pathway.

spectrum is broadly distributed across frequency space, which results in large responses in both the M and P pathways. However, after tracking, the signal energy becomes concentrated at low temporal frequencies, and the P pathway which is more sensitive in the low temporal frequency region, will respond to a greater degree than the M pathway. Figure 9*a* illustrates the response of the M and P pathways to a translating image, found by integrating the spatiotemporal transfer function (equation 8) over lines of constant velocity in frequency space. Figure 9*b*, computed using equations 8 and 9, illustrates the velocity of peak response as a function of eccentricity for the two pathways. At the fovea, the M and P pathways are predicted to have a peak response to images moving at velocities of 7 deg s^{-1} and 1 deg s^{-1}, respectively. However, the velocity of peak response increases with eccentricity, with the peak velocity increasing faster for the M than the P pathway. This change in peak velocity with eccentricity is a result of the increase in receptive field size (or decrease in spatial scale) described by equation 9.

The P pathway is thought to carry information about slowly moving images with a high degree of spatial detail (Schiller *et al.* 1990; Merigan 1989; Merigan *et al.* 1991). Figure 9*a* illustrates that the P pathway will respond better to low velocity images. The P pathway matches the properties of the velocity field component in the tracked region, and will carry the maximum amount of spatiotemporal information about this component. The M pathway is thought to carry information about quickly moving images with a low degree of spatial detail (Schiller *et al.* 1990; Merigan 1989; Merigan & Maunsell 1990). Figure 9*a* illustrates that the M pathway is tuned to higher velocities than the P pathway. However, large image velocities will only rarely arise in the tracked region. When large velocities do arise, it is during tracking errors which occur for unpredictable motion, and for cases such as transparent motion when there are two velocity field components in the same spatial region. Because tracking is generally quite accurate for motion of 'real world' stimuli (Steinman *et al.* 1990), the M pathway can be expected to carry only a small fraction of the velocity field component of image information in the region of tracking (the fovea). However, the stationary component is broadly distributed across spatiotemporal frequency space (figure 2*b*), and contains a significant amount of energy in the region covered by the M pathway. Therefore, in addition to carrying (infrequent) high velocity images, another role for the M pathway at the fovea could be to carry the stationary component of time varying images.

The change of the velocity tuning of the M and P pathways with eccentricity (figure 9*b*) is consistent with the change in the velocity distribution of tracked images with eccentricity. The peak velocity of the two pathways increase with eccentricity, though at different rates, so a larger average velocity and larger range of velocities is covered with increasing eccentricity. This can be compared with the velocity distribution after tracking (figure 4*b*). At the fixation point (fovea), the image has a narrowly distributed velocity distribution and a small temporal bandwidth. With increasing eccentricity, the velocity distribution broadens (figure 4*b*), the average velocity reaching the retina increases (figure 5), and the range of velocities increases. The broader image velocity distribution in the periphery means that information is lost due to temporal blur because of the limited temporal frequency bandwidth of retinal pathways. This decreases the average spatial frequency limit of the peripheral retinal image. Because of this, larger receptive fields can be utilized in the periphery without significant loss of information.

Psychophysical evidence also shows a gradual change in motion perception between the fovea and the periphery of the visual field. The fovea is sensitive to a lower range of velocities than the periphery and essentially becomes blind when this velocity range is exceeded (van de Grind *et al.* 1986; Baker & Braddick 1985). As eccentricity increases, the visual system is better able to discriminate images with a higher average velocity, and over a larger range of velocities (McKee & Nakayama 1984). This is consistent with the change in velocity distribution and average velocity with eccentricity (see figures 4 and 5) which results from tracking.

Hughes (1977) argues that receptor packing matches the change in velocity across the retina for the case of an observer moving through a scene (ego-motion). In some ways, this paper can be viewed as a generalization of Hughes (1977) original arguments, by showing that an eccentricity dependent velocity distribution results for any scene rather than the special case of ego-motion, as long as the observer continually tracks with eye movements. This paper diverges from Hughes by matching the retinal velocity distribution to the velocity sensitivity of the M and P pathways, rather than to the change in receptor packing. However, receptive field size and velocity sensitivity are linked so both arguments are complementary.

5. CONCLUSION

We examined the spatiotemporal spectrum and other attributes of natural time varying images in the context of efficient coding in the early visual system. The image is modeled as a combination of a velocity field component and a stationary component which have markedly different spatiotemporal spectra. Tracking, as implemented with smooth pursuit eye movements, decreases the average velocity and the variability of velocities reaching the fovea (tracked region). The result is a spectrum with minimal temporal bandwidth and variability in the tracked region, but which broadens with increasing eccentricity. Tracking does not affect the stationary component, which remains broadly distributed across temporal frequency space.

An efficient coding strategy will be influenced by tracking because it changes the image spectrum. In the tracked region, the spectrum has minimal temporal bandwidth and variability. This enables efficient coding of the image with fixed time invariant path-

ways as found in the retina. The reduction in temporal bandwidth ensures that minimal information will be lost due to motion blur in the tracked region. The stationary component of time varying images is emphasized in the tracked region, enabling temporal information not attributed to translational motion to be analyzed effectively. Finally, since the average velocity of the image increases with eccentricity from the tracked region, an efficient coding strategy should reflect this change with a corresponding change in velocity tuning with eccentricity.

The results suggest that the M and P pathways are matched to the tracked image. Both the M and P pathways are tuned to low image velocities at the fovea, where the image has consistently low velocities because of tracking. However, the M pathway, with the broader temporal frequency response, will respond better to the temporal changes of the stationary component. The M and P pathways are tuned to higher velocities and a broader range of velocities with increasing eccentricity from the fovea. This is matched to the change of image velocity after tracking, in which both the average velocity and range of velocities increase.

In conclusion, the visual system combines smooth pursuit tracking with specialized pathways and an eccentricity dependent retinal architecture to efficiently code time varying images.

We thank Horace Barlow and Peter Sterling for their many comments and suggestions, and Andrew B. Watson for his aid in collecting the image sequences. This work was supported by AFOSR grant 91-0082 and the NASA graduate student fellowship program.

REFERENCES

Adelson, E.J. & Bergen, J.R. 1985 Spatiotemporal energy models for the perception of motion. *J. opt. Soc. Am.* **A2**, 284–299.

Baker, C.L. & Braddick, O.J. 1985 Eccentricity-dependent scaling of the limits for short-range apparent motion perception. *Vision Res.* **25**, 803–812.

Barlow, H.B. 1961 Possible principles underlying the transformation of sensory messages. In *Sensory communication* (ed. W. A. Rosenblith), pp. 217–234. MIT Press.

Barlow, H.B. 1981 The Ferrier Lecture: critical limiting factors in the design of the eye and visual cortex. *Proc. R. Soc. Lond.* **B212**, 1–34.

Barnes, G.R. & Lawson, J.F. 1989 Head-free pursuit in the human of a visual target moving in a pseudo-random manner. *J. Physiol., Lond.* **410**, 137–155.

Blakemore, C. & Vital-Durand, F. 1986 Organization and post-natal development of the monkey's lateral geniculate nucleus. *J. Physiol., Lond.* **380**, 453–491.

Buchsbaum, G. & Gottschalk, A. 1983 Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proc. R. Soc. Lond.* **B220**, 89–113.

Crook, J.M., Lange-Malecki, B., Lee, B.B. & Valberg, A. 1988 Visual resolution of macaque retinal ganglion cells. *J. Physiol., Lond.* **396**, 205–224.

Derrico, J.B. & Buchsbaum G. 1991 A computational model of spatiochromatic image coding in early vision. *J. visual commun. Image Repres.* **2**, 31–38.

Derrington, A.M. & Lennie, P. 1984 Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *J. Physiol., Lond.* **357**, 219–240.

Eckert, M.P., Buchsbaum, G. & Watson, A.B. 1992 The separability of spatiotemporal spectra of image sequences. *IEEE Trans. Pattern Anal. Machine Intell.* **PAMI-14**, 1210–1213.

Field, D. 1987 Relations between the statistics of natural images and the response properties of cortical cells. *J. opt. Soc. Am.* **A4**, 2379–2394.

Flipse, J.P., Wildt, G.J. v.d., Rodenburg, M., Keemink, C.J. & Knol, P.G.M. 1988 Contrast sensitivity for oscillating sine wave gratings during ocular fixation and pursuit. *Vision Res.* **28**, 819–826.

Girod, B. 1987 The efficiency of motion-compensating prediction for hybrid coding of visual sequences. *IEEE J. Selected Areas Commun.* **5**, 1140–1154.

van de Grind, W.A., Koenderink, J.J. & van Doorn, A.J. 1986 The distribution of human motion detector properties in the monocular visual field. *Vision Res.* **26**, 797–810.

Heeger, D.J. 1987 Model for the extraction of image flow. *J. opt. Soc. Am.* **A4**, 1455–1471.

Hicks, T.P., Lee, B.B. & Vidyasagar, T.R. 1982 The responses of cells in macaque lateral geniculate nucleus to sinusoidal gratings. *J. Physiol., Lond.* **337**, 183–200.

Hughes, A. 1977 The topography of vision in mammals of contrasting life style: comparative optics and retinal organization. In *Handbook of sensory physiology. The visual system in vertebrates* (ed. F. Crescitelli), pp. 613–756. Berlin: Springer-Verlag.

Jain, J.R. & Jain, A.K. 1981 Displacement measurement and its application in interframe image coding. *IEEE Trans. Commun.* **COM-29**, 1799–1808.

Jayant, N.S. & Noll, P. 1984 *Digital coding of waveforms: principles and applications to speech and video*. Englewood Cliffs, New Jersey: Prentice-Hall.

Kaplan, E. & Shapley, R. 1982 X and Y cells in the lateral geniculate nucleus of macaque monkeys. *J. Physiol., Lond.* **330**, 125–143.

Kassam, S.A. & Poor, H.V. 1983 Robust signal processing for communication systems. *IEEE Commun. Mag.* 20–28.

Kassam, S.A. & Poor, H.V. 1985 Robust techniques for signal processing. *Proc. IEEE* **73**, 433–481.

Kelly, D.H. 1979 Motion and vision II: stabilized spatio-temporal threshold surface. *J. opt. Soc. Am.* **A69**, 1340–1349.

Laughlin, S. 1983 Matching coding to scenes to enhance efficiency. In *Physical and biological processing of images* (ed. O. J. Braddick & A. Sleigh), pp. 42–52. Berlin: Springer-Verlag.

Lee, B.B., Martin, P.R. & Valberg, A. 1989*a* Amplitude and phase of responses of macaque retinal ganglion cells to flickering stimuli. *J. Physiol., Lond.* **414**, 245–263.

Lee, B.B., Martin, P.R. & Valberg, A. 1989*b* Sensitivity of macaque retinal ganglion cells to chromatic and luminance flicker. *J. Physiol., Lond.* **414**, 223–243.

Lisberger, S.G., Morris, E.J. & Tychsen, L. 1987 Visual motion processing and sensory-motor integration for smooth pursuit eye movements. *A. Rev. Neurosci.* **10**, 97–129.

Marrocco, R.T., McClurkin, J.W. & Young, R.A. 1982 Spatial summation and conduction latency classification of cells of the lateral geniculate nucleus of macaques. *J. Neurosci.* **2**, 1275–1291.

McKee, S.P. & Nakayama, K. 1984 The detection of motion in the peripheral visual field. *Vision Res.* **24**, 25–32.

Merigan, W.G. 1986 Spatio-temporal vision of macaques with severe loss of P retinal ganglion cells. *Vision Res.* **26**, 1751–1761.

Merigan, W.H. 1989 Assessing the role of parallel pathways in primates. In *Seeing contours and colour* (ed. J. J. Kulikowski, C. M. Dickinson & I. J. Murray). Pergamon Press.

Merigan, W.H. & Maunsell, J.H.R. 1990 Macaque vision after magnocellular lateral geniculate lesions. *Vis. Neurosci.* **5**, 347–352.

Merigan, W.H., Katz, L.M. & Maunsell, J.H.R. 1991 The effects of parvocellular lateral geniculate lesions on the acuity and contrast sensitivity of macaque monkeys. *J. Neurosci.* **11**, 994–1001.

Miller, J.W. & Ludvigh, E. 1962 The effect of relative motion on visual acuity. *Surv. Opthalm.* **7**, 83–116.

Murphy, B.J. 1978 Pattern thresholds for moving and stationary gratings during smooth eye movement. *Vision Res.* **18**, 521–530.

Pentland, A. 1991 Photometric motion. *IEEE Pattern Anal. Machine Intell.* **13**, 879–890.

Purpura, K., Tranchina, D., Kaplan, E. & Shapley, R.M. 1990 Light adaptation in the primate retina: analysis of changes in gain and dynamics of monkey retinal ganglion cells. *Vis. Neurosci.* **4**, 75–93.

Rohaly, A.M. & Buchsbaum, G. 1988 Inference of global spatiochromatic mechanisms from contrast sensitivity functions. *J. opt. Soc. Am.* **A5**, 572–576.

Rohaly, A.M. 1988 A global multidimensional model of human visual contrast sensitivity. Ph.D. thesis. Department of Bioengineering, University of Pennsylvania.

van Santen, J.P.H. & Sperling, G. 1985 Elaborated Reichardt detectors. *J. opt. Soc. Am. A* **2**, 300–321.

Sakitt, B. & Barlow, H.B. 1982 A model for the economical coding of the visual images in cerebral cortex. *Biol. Cybern.* **43**, 97–108

Schiller, P.H., Logothetis, N.K. & Charles, E.R. 1990 role of the color-opponent and broad-band channels in vision. *Vis. Neurosci.* **5**, 321–346.

Shapley, R.M. & Perry, V.H 1986 Cat and monkey retinal ganglion cells and their visual functional roles. *Trends Neurosci.* 229–235.

Snyder, A.W., Laughlin, S.B. & Stavenga, D.G. 1977 Information capacity of eyes. *Vision Res.* **17**, 1163–1175.

Srinivisan, M.V., Laughlin, S.B. & Dubs, A. 1982 Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. Lond.* **B216**, 427–459.

Stark, L., Vossius, G. & Young, L.R. 1962 Predictive control of eye tracking movements. *IRE Trans. Human Factors Electr.* 52–57.

Steinman, R.M., Kowler, E. & Collewijn, H. 1990 New directions for oculomotor research. *Vision Res.* **30**, 1845–1864.

Tsukamoto, Y., Smith, R.G. & Sterling, P. 1990 'Collective coding' of correlated cone signals in the retinal ganglion cells. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1860–1864.

Watson, A.B. 1983 A look at motion in the frequency domain. NASA Tech. Report Tech. Memo. 84352.

Watson, A.B. & Ahumada, A.J. 1985 Model of human visual motion sensing. *J. opt. Soc. Am.* **A2**, 322–341.

Watson, A.B. 1990 Perceptual-components architecture for digital video. *J. opt. Soc. Am.* **A7**, 1943–1954.

Received 4 March 1992; accepted 19 October 1992

Reprinted from Journal of the Optical Society of America A

Effect of tracking strategies on the velocity structure of two-dimensional image sequences

Michael P. Eckert and Gershon Buchsbaum

University of Pennsylvania, School of Engineering and Applied Science, Department of Bioengineering, 220 South 33rd Street, Philadelphia, Pennsylvania 19104-6315

Received November 23, 1992; accepted February 10, 1993

We investigate the effect of different tracking strategies, such as local and full-field tracking, on the mean and variance of the image velocity field. We show that while local tracking reduces the velocity variability in an eccentricity-dependent manner, full-field tracking reduces velocity variability equally across the image. We test our predictions with digitized image sequences.

INTRODUCTION

It is well known that tracking actively modifies the spatiotemporal structure of the visual scene; it shifts image velocity and position, thus making the tracked image markedly different from the original image.[1–6] In this Communication we examine and formalize how different tracking strategies affect the local and global velocity structures of image sequences. To describe the local velocity structure, we use the velocity field, which assigns a velocity vector to the image at each point in space and time and is specific to each scene. For a more global description of image velocity, we calculate the variance of the velocity field, sampled through time. We focus on two strategies of tracking: local tracking, in which a small spatial region in the original image is pursued, and full-field tracking, in which the average image velocity is tracked. These two tracking strategies reduce the variance of the velocity distribution in markedly different ways. Local tracking greatly reduces the variance at the point of tracking, but the variance increases with eccentricity away from the point of tracking in a manner dependent on the spatial correlation of the velocity field. Full-field tracking, on the other hand, reduces the variance equally across the entire spatial extent of the image. We illustrate the analysis with examples by using image sequences.

DEFINITION OF TRACKING

We define the velocity of tracking as a space average of image velocity, in which different strategies of tracking are associated with the spatial extent over which image velocity is averaged:

$$\mathbf{v}_t(t) = \int \mathbf{v}(\mathbf{u}, t) g(\mathbf{u}) d\mathbf{u}, \tag{1}$$

where $\mathbf{v}_t(t)$ is the velocity of tracking, $\mathbf{v}(\mathbf{u}, t)$ is the velocity field of the original image, $g(\mathbf{u})$ limits the spatial extent over which velocity is averaged, and $\mathbf{u}$ is space in coordinates of the image plane. The integration is over the spatial extent of the image.

For the purpose of analysis, we consider only two cases of tracking. The first, local tracking, sets the tracking velocity to the image velocity in a small spatial region of the image plane. In the limit, this represents tracking of a single point [$g(\mathbf{u}) \rightarrow \delta(\mathbf{u} - \mathbf{u}_0)$]. The second case, full-field tracking, sets the tracking velocity to the average velocity in the image plane [$g(\mathbf{u}) \rightarrow 1$].

VELOCITY FIELD OF THE TRACKED IMAGE

The velocity field assigns a velocity vector to each point of the image plane to represent intensity variations resulting from geometric motion in the world. Types of motion that are adequately described by the velocity field include affine transformations such as translation, rotation, and dilation that are commonly found with perspective projections of three-dimensional motion.[7,8]

We write the velocity field as the sum of two components,

$$\mathbf{v}(\mathbf{u}, t) = \mathbf{v}_d(\mathbf{u}, t) + \mathbf{v}_m(t), \tag{2}$$

where $\mathbf{v}_m(t)$ denotes the time-varying mean velocity and $\mathbf{v}_d(\mathbf{u}, t)$ denotes differential image velocities that vary with both space and time. Each of these components can be associated with common physical sources of motion. Changes in the time-varying mean velocity occur during camera pans (or head and body rotation in the biological case). $\mathbf{v}_d(\mathbf{u}, t)$ represents differential velocities across the field of view, such as motion of objects or linear motion of the camera through the world (ego motion).

Tracking introduces a single time-varying vector-velocity term to the velocity field. After tracking, the velocity field is simply the difference between the velocity field of the original image and the velocity of tracking:

$$\mathbf{v}_r(\mathbf{u}, t) = \mathbf{v}(\mathbf{u}, t) - \mathbf{v}_t(t), \tag{3}$$

0740-3232/93/071582-04$06.00

where $\mathbf{v}_r(\mathbf{u}, t)$ is the tracked velocity field. This relationship is valid for any type of tracking strategy, including both full-field and local tracking.

We consider the effect that the two cases of tracking have on the velocity field. Local tracking, in the limit, consists of tracking a single point, $\mathbf{v}_t(t) = \mathbf{v}(\mathbf{u}_0, t)$. And using Eq. (2) gives

$$\mathbf{v}_t(t) = \mathbf{v}_d(\mathbf{u}_0, t) + \mathbf{v}_m(t). \tag{4}$$

Substituting Eqs. (2) and (4) into Eq. (3) shows that local tracking removes the mean velocity, leaving a tracked velocity field of

$$\mathbf{v}_r(\mathbf{u}, t) = \mathbf{v}_d(\mathbf{u}, t) - \mathbf{v}_d(\mathbf{u}_0, t). \tag{5}$$

Full-field tracking, with the definition from Eq. (1), amounts to setting the velocity of tracking to the time-varying mean image velocity, i.e., $\mathbf{v}_t(t) = \mathbf{v}_m(t)$. Thus the velocity field after full-field tracking is simply that of the differential velocity term,

$$\mathbf{v}_r(\mathbf{u}, t) = \mathbf{v}_d(\mathbf{u}, t). \tag{6}$$

VARIANCE OF THE VELOCITY FIELD AFTER TRACKING

The velocity field provides an instantaneous and locally specific measure of image velocity. A more global measure is produced by calculating the mean and variance of the original and tracked velocity fields. For this purpose, we consider the velocity field of an image sequence to be a realization of a three-dimensional (two-dimensional space, one-dimensional time) vector random field. The mean and variance are then calculated by sampling the velocity field across space, through time, or both.[9] In this Communication, velocity samples are collected through time at every point in space. Conceptually, these samples form a velocity distribution or density function and represent a time-averaged measure of image velocity at that point in space.[10] Since velocity is a vector quantity, the mean and variance of the horizontal and vertical velocity components are calculated separately and are considered independent and uncorrelated. We assume that $\mathbf{v}_0(\mathbf{u}, t)$ and $\mathbf{v}_m(t)$ have a zero temporal mean. The basis for this assumption is that, on average, there will be no long-term fixed mean-velocity bias in image sequences.

After local tracking, the mean velocity will be zero, but the velocity variance will depend on distance from the point of tracking:

$$\sigma_r(\mathbf{u} - \mathbf{u}_0)^2 = \sigma_d{}^2(\mathbf{u}) + \sigma_d{}^2(\mathbf{u}_0) - 2B_d(\mathbf{u} - \mathbf{u}_0), \tag{7}$$

where $\sigma_r(\mathbf{u} - \mathbf{u}_0)^2$ is the velocity variance of the locally tracked image as a function of distance from the point of tracking, $\sigma_0{}^2(\mathbf{u})$ is the variance of the differential velocity term, $\sigma_d{}^2(\mathbf{u}_0)$ is the variance of velocity at point $\mathbf{u}_0$, and $B_d(\mathbf{u} - \mathbf{u}_0)$ is the spatial autocorrelation function of the differential velocity term. The key point to recognize here is that the variance approaches zero at the point of tracking and will increase regularly with distance from the tracked point for a monotonically decreasing velocity-field autocorrelation function. At the largest eccentricities, the velocity-field autocorrelation approaches zero, so the variance is simply the sum of the first two terms.

Full-field tracking removes the time-varying mean from the original image. As a result, the full-field tracked image will have a zero mean and a variance equal to the variance of the differential velocity term,

$$\sigma_r(\mathbf{u}) = \sigma_d{}^2(\mathbf{u}). \tag{8}$$

The variance in Eq. (8) depends on the object velocity and is not a function of the spatial correlation of the velocity field.

Figure 1 illustrates the effect of local and full-field tracking, assuming constant variance for the differential velocity term and a spatial correlation of the velocity field that falls off exponentially with space. The variance of the velocity field after local tracking is greatly reduced in the region of tracking but increases with larger eccentricities. The reason for this is that the velocity field of the image is correlated with tracking velocity at the point of tracking but is almost completely uncorrelated at large distances from that point. Far from the point of tracking, $|\mathbf{u}| \gg 0$, the variance of the velocity approaches twice the variance of the object-velocity term in the original image. Between these endpoints, the variance of the velocity field depends on the spatial correlation of the velocity field. Full-field tracking removes the mean velocity, $\mathbf{v}_m(t)$, causing a downward shift of the variance from σ^2 for the original image to $\sigma_0{}^2$ for the tracked image. That is, the reduction in variance for full-field tracking is not dependent on eccentricity and amounts to removal of the variance contributed by the mean-velocity term, $\mathbf{v}_m(t)$. An implicit assumption is that the correlation between the differential velocity term and the time-varying mean velocity is small. In limiting cases, such as when a single object moves across a fixed background, the two terms are obviously correlated. However, the examples below, with the use of natural image sequences, suggest that the as-

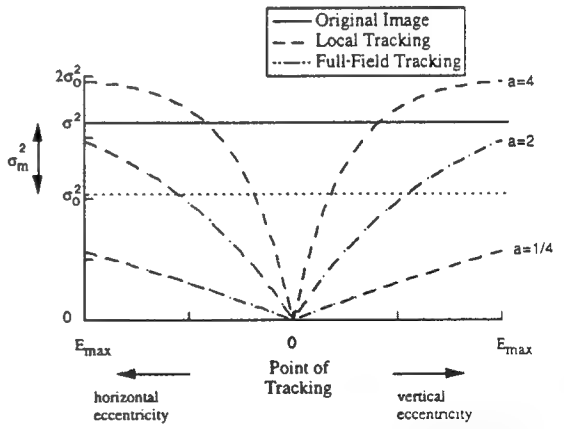


Fig. 1. Predicted effect of local and full-field tracking on the variance of the velocity field as a function of eccentricity from the point of tracking. We use a spatial autocorrelation function for the differential velocity field of the form $B_d(\mathbf{u}) = \exp(-a|\mathbf{u}|/E_{max})$. The dashed curves represent the variance of local tracking for different spatial autocorrelation functions of the velocity field when $a = (1/4, 2, 4)$. The bottom curve ($a = 1/4$) illustrates the case in which the velocity field is highly correlated across space. The top curve ($a = 4$) represents the case in which the velocity field is relatively uncorrelated across space. The endpoint of the top curve asymptotes at twice the variance of the full-field tracked image. Full-field tracking reduces the variance from σ^2 to $\sigma_d{}^2$ and is independent of the correlation of the velocity field.

Table 1. Description of Image Sequences

Sequence Number	Sequence Description
IJ10833	Jungle scene with some three-dimensional object motion and camera motion.
IJ12426	Man walking. Some camera motion to keep man centered in visual field. Result is considerable background motion.
IJ01300	Man talking while moving head occasionally. No camera motion. Some motion in the background.

sumption is justified. For another limiting case, that of purely full-field motion, in which there is only a single translational velocity in the image, the differential velocity term, $\mathbf{v}_d(\mathbf{u}, t)$, is zero, and local tracking decreases the variance at all eccentricities. For this case full-field and local tracking produce the same effect.

EFFECT OF TRACKING ON IMAGE SEQUENCES

We simulated local and full-field tracking on three digitized image sequences and calculated the velocity field of the original image, the image after local tracking, and the image after full-field tracking. The sequences (256 × 256 × 64 pixels at 8 bits/pixel, 30 frames/s with no scene cuts)[11] were taken from a video disk that contained scenes from movies (see Table 1). Local tracking was initialized by selecting a 24 × 24 pixel block in the first frame of the sequence. This region was then tracked by minimizing the squared difference between 24 × 24 pixel blocks in each pair of sequential frames in the sequence.[12] The frames of the sequence were shifted so as to maintain the tracked object in the same spatial location for the entire sequence. This algorithm was effective because the scenes generally consisted of rigid, moving objects. Full-field tracking was implemented by shifting each frame in the sequence by the mean velocity of the entire image rather than by the velocity in a small region. With full-field tracking, no region was guaranteed to remain static throughout the sequence. The velocity field was calculated in a sparse array (every 16 pixels) by finding the displacement that minimized the squared difference between blocks in two sequential frames of the sequence.[12] Thus the local velocity structure of each sequence was described by a 16 × 16 × 64 cube of velocity vectors. The original velocity field was calculated from the original image, and the local and the full-field tracked velocity fields were calculated from the scene after tracking. The variance of the original, the full-field, and the local tracked velocity fields were formed by sampling through time at each spatial location. The variance of each velocity distribution was then calculated at selected vertical and horizontal eccentricities from the tracked point.

Figure 2 illustrates the variance of the velocity field for the original image (solid curves), the local tracked image (dashed curves), and the full-field tracked image (dotted curves). The variance represents samples collected from a horizontal or a vertical eccentric point from the point of tracking. The different magnitudes of the variance from sequence to sequence reflect the different levels of motion activity in each sequence. Figure 2 also shows that the variance of the original velocity field can vary significantly across space. This is not surprising since a short, 2-s sequence will usually have different levels of motion activity in different areas of the scene. However, there was no systematic bias in motion activity across the scene in the different sequences.

Local tracking removes the variance of image velocity in the region of tracking but at the expense of increasing the variance at large eccentricitiès. This feature is consistent for all sequences, independent of the level of motion activity in the sequence. The rate at which the variance changes with eccentricity depends on the spatial correlation of the velocity field in each image. At large eccentricities, the variance of the local tracked image is greater than the variance of the full-field tracked image. This is expected because of the diminishing correlation of the velocity field between the tracked point and points at large

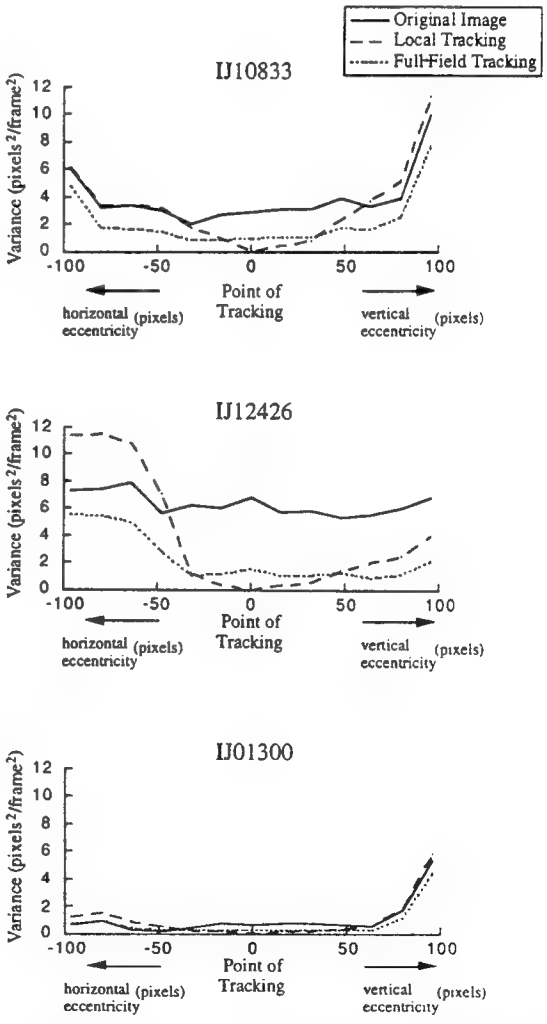


Fig. 2. Variance of the velocity field of three sequences as a function of distance from the tracked point. The variances of the horizontal and the vertical velocity components were added together. The curves represent the variance of the original image (solid curves), the local tracked image (dashed curves), and the full-field tracked image (dotted curves). Full-field tracking reduces the variance at all eccentricities, but local tracking reduces the variance primarily in the region of tracking.

eccentricities. As expected, the reduction in variance after full-field tracking is well described by a downward shift of the variance of the original image. Differences among the variance curves that are not accounted for by the downward shift can be attributed to the correlation between the differential velocity field and the time-varying mean velocity and are small for all sequences.

DISCUSSION

Local or full-field tracking almost always reduces the variance of the velocity field. Local tracking ensures that one area of the visual field will possess low velocities but at the expense of increasing image velocity at large eccentricities from the point of tracking. Full-field tracking reduces velocities across the entire field of view, but no area is assured of consistently low velocities. If velocity magnitude of the tracked image is averaged across the extent of the image, with equal weight given to all areas, then the full-field tracked image will have a lower space-averaged variability than the local tracked image. This effect can be seen by inspection of Fig. 2. However, if the primary constraint is to reduce image velocity maximally in a small area, then local tracking is obviously more suitable. There are also other advantages of local tracking, which include regularizing high-level visual tasks such as structure from motion, wayfinding, and shape from shading.[6] It remains to be seen whether full-field tracking can provide similar benefits.

An original motivation of this study was the significance of biological tracking strategies as implemented by various species using head and eye movements. Local tracking is a reasonable approximation of smooth-pursuit eye movements that are found in some vertebrate species, and full-field tracking is a good approximation of optokinetic and vestibulo-ocular eye movements that are found in most, if not all, vertebrate species.[13–17] As illustrated in this communication, different tracking strategies result in different degrees of velocity variability across the visual field. This finding suggests that biological velocity sensitivity at various eccentricities in a given species should reflect the particular tracking strategy most common to that species.[10] Determining whether this is so requires further comparative research into tracking capabilities and eccentricity-dependent velocity sensitivity in different species.

ACKNOWLEDGMENT

This research was supported by U.S. Air Force Office of Scientific Research grant 91-0082.

REFERENCES AND NOTES

1. B. J. Murphy, "Pattern thresholds for moving and stationary gratings during smooth eye movement," Vision Res. **18,** 521–530 (1978).
2. J. P. Flipse, G. J. van Wildt, M. Rodenburg, and P. G. M. Knol, "Contrast sensitivity for oscillating sine wave gratings during ocular fixation and pursuit," Vision Res. **28,** 819–826 (1988).
3. B. Girod, "Eye movements and coding of video sequences," in *Visual Communications and Image Processing, '88: Third in a Series,* T. R. Hsing, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **1001,** 398–405 (1988).
4. J. H. D. M. Westerink and C. Teunissen, "Perceived sharpness in moving images," in *Human Vision and Electronic Imaging: Models, Methods, and Applications,* J. P. Allebach and B. E. Rogowitz, eds., Proc. Soc. Photo-Opt. Instrum. Eng. **1249,** 78–87 (1990).
5. W. H. Warren and D. J. Hannon, "Eye movements and optical flow," J. Opt. Soc. Am. **7,** 160–169 (1990).
6. Y. Aloimonos and A. Rosenfeld, "Computer vision," Science **253,** 1181–1324 (1991).
7. P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," Int. J. Comput. Vision **2,** 283–310 (1990).
8. D. Fleet and A. D. Jepson, "Computation of component image velocity from local phase information," Int. J. Comput. Vision **5,** 77–104 (1990).
9. E. Vanmarcke, *Random Fields: Analysis and Synthesis* (MIT Press, Cambridge, Mass., 1983).
10. M. P. Eckert and G. Buchsbaum, "Efficient coding of natural time varying images in the early visual system," Phil. Trans. R. Soc. London (to be published).
11. These digitized images are available from the authors.
12. J. R. Jain and A. K. Jain, "Displacement measurement and its application in interframe image coding," IEEE Trans. Commun. **COM-29,** 1799–1808 (1981).
13. G. L. Walls, "The evolutionary history of eye movements," Vision Res. **2,** 69–80 (1962).
14. P. E. Hallett, "Eye movements," in *Handbook of Perception and Human Performance,* K. R. Boff, L. Kaufman, and J. P. Thomas, eds. (Wiley, New York, 1986), Chap. 10.
15. D. A. Robinson, "The use of control system analysis in the neurophysiology of eye movements," Ann. Rev. Neurosci. **4,** 463–503 (1981).
16. H. Collewijn, "The optokinetic system," in *Models of Oculomotor Behavior and Control,* B. L. Zuber, ed. (CRC, Boca Raton, Fla., 1981).
17. S. G. Lisberger, E. J. Morris, and L. Tychsen, "Visual motion processing and sensory-motor integration for smooth pursuit eye movements," Ann. Rev. Neurosci. **10,** 97–129 (1987).

Appendix C

Separability of Spatiotemporal Spectra of Image Sequences

Michael P. Eckert, Gershon Buchsbaum, and Andrew B. Watson

Abstract— **We calculated the spatiotemporal power spectrum of 14 image sequences in order to determine the degree to which the spectra are separable in space and time and to assess the validity of the commonly used exponential correlation model found in the literature. We expand the spectrum by a singular value decomposition into a sum of separable terms and define an index of spatiotemporal separability as the fraction of the signal energy that can be represented by the first (largest) separable term. All spectra were found to be highly separable with an index of separability above 0.98. The power spectra of the sequences were well fit by a separable model of the form**

$$P(k, f) = \frac{ab/(4\pi^3)}{((a/2\pi)^2 + k^2)^{3/2}((b/2\pi)^2 + f^2)}$$

where k is radial spatial frequency, f is temporal frequency, and a,b are spatial and temporal model parameters that determine the effective spatiotemporal bandwidth of the signal. This power spectrum model corresponds to a product of exponential autocorrelation functions separable in space and time.

I. Introduction

The statistics of images and image sequences have been extensively studied for image coding and compression applications [1], [2] as well as for the development of models of biological image processing [3], [4]. An exponential autocorrelation function has been shown to be a good model for temporal frame-to-frame correlations of image sequences, e.g., [5]–[8], and for spatial correlations within each frame, e.g., [2], [3], [9].

This paper focuses on the separability of the spatiotemporal statistics of image sequences and on the validity of using a separable exponential autocorrelation model for the spatiotemporal statistics. The autocorrelation function is uniquely related to the power spectrum via a Fourier transform, and either is valid as a description of the statistics.

Manuscript received November 7, 1990; revised March 6, 1992. This work was supported by the NASA Graduate Fellowship Program and by grants NSF 8351637 and AFOSR 91-0082. Recommended for acceptance by Associate Editor N. Ahuja.

M. P. Eckert and G. Buchsbaum are with the Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA 19104-6392.

A. B. Watson is with the NASA Ames Research Center, Moffet Field, CA 94035-1000.

IEEE Log Number 9204244.

The spectra of 14 image sequences were calculated. The sequences represented a small ensemble of possible motion activity. The sequences were selected for a range of motion activity. For example, a fast camera pan represents the maximum image motion activity, and a small moving object with a static background represents the least activity. Sequences with motion activity between these extremes had slight camera motion and some object motion.

II. Calculation of Image Statistics

We collected 14 image sequences (256 × 256× 64 @ 8 b/pixel, 30 frames/s with no scene cuts) from a video disc that contained scenes from a broadcast TV source. Each frame was originally sampled at 512 × 512 pixels/screen, but adjacent pixels were averaged, and the image was subsampled to 256 × 256 pixels/screen. The sample mean of each sequence was removed to reduce low-frequency bias in the calculations.

The sample power spectrum $P(k_1, k_2, f)$ of each sequence $x(n_1, n_2, t)$ is the squared magnitude of the discrete Fourier transform calculated as

$$P(k_1, k_2, f) = \frac{1}{256 \cdot 256 \cdot 64} \cdot \left| \sum_{n_1=0}^{255} \sum_{n_2=0}^{255} \sum_{t=0}^{63} x(n_1, n_2, t) e^{-j2\pi(k_1 n_1 + k_2 n_2 + ft)} \right|^2 \quad (1)$$

where k_1,k_2 are spatial frequencies, f is temporal frequency, n_1,n_2 are spatial locations, and t is time measured in frame number.

We converted the two spatial frequency dimensions k_1 and k_2 into one radial frequency dimension k by averaging in 32 annuli around the spatial frequency origin as illustrated in Fig. 1. In this manner, the spatial frequency range of 0–127 cycles/screen of k_1 and k_2 is represented by 32 annuli in bands of 4 cycles/screen. Averaging the spatial spectra in annuli is equivalent to assuming a circularly symmetric spatial autocorrelation function. This autocorrelation function is not separable in the two spatial dimensions but is considered a better fit than the corresponding separable autocorrelation function for most images [9].

The average magnitude of the power spectrum in each annulus can be obtained by summing over the power spectrum $P(k_1, k_2, f)$ in the annulus indexed by k and normalizing by the number of sample

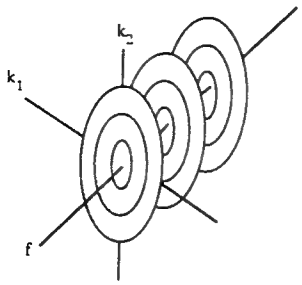


Fig. 1. Conversion from two dimensions of spatial frequency to one dimension of radial spatial frequency is done by averaging the spectrum in annuli around the spatial frequency origin.

points $A(k)$ within each annulus

$$P(k,f) = \frac{1}{A(k)} \sum_{k_1^2+k_2^2 \geq k^2}^{k_1^2+k_2^2<(k+4)^2} P(k_1,k_2,f) \quad k = 0,4,8,\ldots 124 \tag{2}$$

where

$$A(k) = \sum_{k_1^2+k_2^2 \geq k^2}^{k_1^2+k_2^2<(k+4)^2} 1 \quad k = 0.4.8.\ldots 124. \tag{3}$$

The resulting 14 sample spectra were described in terms of a 33 (temporal frequency) × 32 (spatial frequency) matrix P with the spatial frequency axis ranging from 0–127 cycles/screen in steps representing bands of 4 cycles/screen and the temporal frequency axis ranging from 0–15 Hz in steps of 15/32 Hz each.

III. Models of Space-Time Statistics

The most commonly used statistical model for intraframe and frame-to-frame correlations is an exponential correlation model in both space and time

$$R(\nu) = e^{-a|\nu|} \tag{4}$$

$$R(\tau) = e^{-b|\tau|} \tag{5}$$

where ν represents a 2-D spatial lag, τ represents temporal lags, and a,b are spatial and temporal parameters. A separable formulation for the spatiotemporal correlation of image sequences is found as a product of (4) and (5). An equivalent description of the statistics is the power spectrum, which for the exponential correlation function of (4) and (5) would be

$$S(k) = \frac{a/(2\pi)}{\left((a/(2\pi))^2 + k^2\right)^{3/2}} \qquad k \geq 0 \tag{6}$$

$$T(f) = \frac{b/(2\pi^2)}{(b/(2\pi))^2 + f^2} \qquad -\infty < f < \infty \tag{7}$$

where k is radial spatial frequency, f is temporal frequency, a is a spatial parameter with units of cycles/screen, and b is a temporal parameter with units of Hertz. The parameters a and b describe the effective spatial and temporal bandwidth of the signal. A spatial power spectrum (6) has 85% of its power in the frequency band $k \leq a$. The temporal power spectrum (7) has 90% of its power in the band $f \leq |b|$. A separable spatiotemporal power spectrum is formed as the product of (6) and (7).

$$P(k,f) = \frac{ab/(4\pi^3)}{\left((a/2\pi)^2 + k^2\right)^{3/2}\left((b/2\pi)^2 + f^2\right)}. \tag{8}$$

IV. Singular Value Decomposition and Index of Separability

A space-time separable spectrum is modeled as the product of a spatial and temporal spectrum (as in (8)). In this section, we define an index of separability for an arbitrary spectrum $P(k,f)$ based on a singular value decomposition.

Any $m \times n$ matrix $\boldsymbol{D}$ with $m \geq n$ may be expanded into a sum of terms by a singular value decomposition [10], [11]

$$\boldsymbol{D} = \sum_{i=1}^{n} \sqrt{\gamma_i}\boldsymbol{v}_i\boldsymbol{u}_i \tag{9}$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n$ are the real nonnegative eigenvalues of the nth-order symmetric matrix $\boldsymbol{S} = \boldsymbol{D}^T\boldsymbol{D}$. $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots \boldsymbol{u}_n$ are normalized, orthogonal row eigenvectors associated with the corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n$ of $\boldsymbol{S}$. $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots \boldsymbol{v}_n$ are normalized, orthogonal column eigenvectors associated with the corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n$ of the mth-order symmetric matrix $\boldsymbol{Q} = \boldsymbol{D}\boldsymbol{D}^T$, where $\boldsymbol{Q}$ can have a maximum of n nonzero eigenvalues that are the same as those of $\boldsymbol{S}$. In the case of duplicate eigenvalues, an orthonormal combination of eigenvalues can be selected.

Approximating $\boldsymbol{D}$ by the first term of the decomposition

$$\boldsymbol{D}' = \sqrt{\gamma_1}\boldsymbol{v}_1\boldsymbol{u}_1 \tag{10}$$

gives the minimum mean squared error separable approximation to $\boldsymbol{D}$, where the mean squared error is

$$e = \sum_{i=1}^{n}\sum_{j=1}^{m} \left(d - d'\right)_{ij}^2 \tag{11}$$

where d_{ij} and d'_{ij} are the elements of $\boldsymbol{D}$ and $\boldsymbol{D}'$, respectively. Noting that

$$\sum_{i=1}^{n}\sum_{j=1}^{m} d_{ij}^2 = \sum_{i=1}^{n} \gamma_i \tag{12}$$

and

$$\sum_{i=1}^{n}\sum_{j=1}^{m} d_{ij}'^2 = \gamma_1$$

the mean square error between the approximate matrix $\boldsymbol{D}'$ and the true matrix $\boldsymbol{D}$ is determined by the eigenvalues as

$$e = \gamma_2 + \gamma_3 + \ldots \gamma_n. \tag{13}$$

We define an index of separability α as the relative energy share of $\boldsymbol{D}'$

$$\alpha = \frac{\gamma_1}{\gamma_1 + \gamma_2 + \ldots \gamma_n}. \tag{14}$$

Since $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n \geq 0$, α will range from $1/n$ for the most inseparable spectrum to 1 for a completely separable spectrum. The eigenvalues represent the energy carried by each term of the expansion in (9). The index of separability α is simply the fraction of the total energy carried by the first and largest term in the expansion, which is the term that constitutes the best separable approximation.

We applied the singular value decomposition to the spatiotemporal spectra by considering each spectrum as a matrix $\boldsymbol{P}$ of dimension 33×32. As shown in (9), $\boldsymbol{P}$ can be expanded as

$$\boldsymbol{P} = \sum_{i=1}^{32} \sqrt{\gamma_i}\boldsymbol{t}_i\boldsymbol{s}_i \tag{15}$$

where $\boldsymbol{s}_i$ are now orthonormal row vectors representing spatial spectra, and $\boldsymbol{t}_i$ are orthonormal column vectors representing temporal

TABLE I
DESCRIPTION OF IMAGE SEQUENCES AND RESULTS OF CALCULATIONS.

	Sequence Number	Motion Type	Index of Separability α	Spatial Parameter a	Temporal Parameter b	mse (%)
1	IJ01300	1,a	0.999	14.33	0.59	0.01
2	IJ04454	1,b	0.999	7.54	0.51	0.04
3	IJ10833	2,a	0.993	9.45	1.08	0.09
4	IJ10897	2,a	0.995	9.42	1.30	0.07
5	IJ11907	1,c	0.999	6.91	3.50	4.70
6	IJ12100	1,a	0.999	15.80	0.41	0.03
7	IJ12164	1,b	0.999	13.85	0.92	0.06
8	IJ12426	2,b	0.998	8.10	0.92	0.04
9	IJ14461	3,a	0.998	6.00	4.30	0.70
10	IJ15300	3,b	0.997	8.93	2.99	4.10
11	IJ17830	1,c	0.982	12.30	2.32	0.40
12	IJ07860	1,c	0.993	11.50	1.85	0.60
13	IJ33960	1,a	0.999	10.20	0.24	0.005
14	IJ30229	1,b	0.999	12.40	0.85	0.06

α : Index of separability, unitless
a : Spatial parameter, cycles/screen
b : Temporal parameter, Hertz
mse : The mean squared error between the actual spectrum and the model with the parameters a,b of Eq. 8. The mse is expressed as the percentage of the average power of the sequence.

1. No camera motion
2. Some camera motion
3. Much camera motion
a. Little object motion
b. Some object motion
c. Much object motion

spectra in each term of the sum. A separable approximation of the form

$$P' = \sqrt{\gamma_1}\, t_1 s_1 \tag{16}$$

exists where s_1 and t_1 represent the spatial and temporal components of the separable approximation. The normalized energy share of this term is α, which is the index of separability. Examination of α for the spatiotemporal spectra of the 14 image sequences (Table I) shows that for 13 out of the 14 sequences, $\alpha > 0.993$, which constitutes a high degree of separability [10]. Although the separability was low for one sequence,($\alpha = 0.982$). This suggests that a space-time separable model such as (8) may adequately describe the spatiotemporal spectrum of image sequences since the assumption of separability is valid. The extraction of nearly all the energy with the separable term is also significant for perceptual reasons since small fractions of image energy can markedly affect the perception of some images [12].

V. CALCULATION OF MODEL PARAMETERS

Since the spatiotemporal spectra of the image sequence P are all highly separable, we need only determine whether the model of (8) adequately characterizes the frequency distribution of the spectra and find the spatial and temporal parameters a and b. This will determine whether the commonly used model defined by a separable exponential autocorrelation in space and time is satisfactory.

We find the model parameters a and b by minimizing the mean squared error between the actual signal spectra P of (2) and the analytical separable model of (8).

$$\min\left[(P - P(k,f))^2\right]. \tag{17}$$

The optimal parameters a,b for each of the sequences were calculated using the Nelder-Meade simplex algorithm [13]. The mean squared error between the analytical separable model (8) and the true spectrum, which was expressed as a percentage of the average squared power of the spectrum, is small ($0.03\% < \text{mse} < 4.7\%$) and is given in Table I. The parameters a and b determine the effective bandwidth for the spatiotemporal power spectrum. Fig. 2 illustrates the relationship between the parameters a and b for all 14 sequences, and thus, the simultaneous spatial and temporal bandwidths. All of the pairs of a and b are located within a well-defined range for this ensemble such that no sequence contains both high spatial and high temporal frequencies.

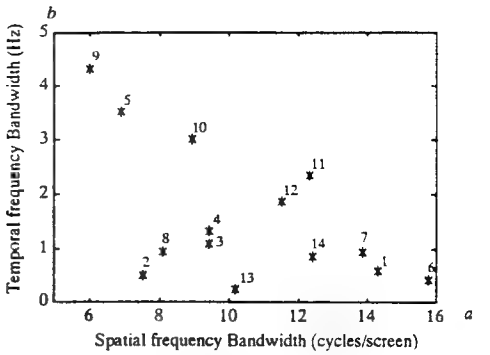


Fig. 2. Scatter plot of the parameters a and b for all sequences. The parameters a and b are measures of the effective spatial and temporal bandwidths of the signal spectrum. No spectrum had both a large spatial and large temporal bandwidth within the spatial and temporal frequency spans of the sequences.

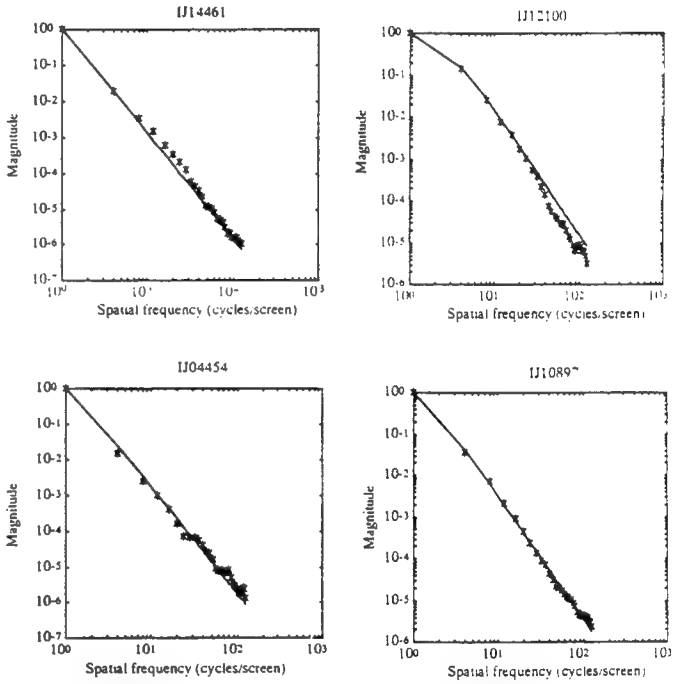


Fig. 3. Magnitude of the spatial component of the spectrum derived by the singular value decomposition (stars) compared with the analytical model (solid line). (Note different ordinate scales.)

The separable kernel in the model of (8) is based on theoretical considerations, mainly, statistical properties of Markov processes as models for image signals. It is interesting to investigate how this theoretical separable model captures the functional shape of the spectra in spatial and temporal frequency compared with the empirically derived separable kernels derived by the singular value decomposition. The empirically derived kernels are not constrained by a predetermined functional shape as is the theoretical model. We compare the spatial and temporal components of the analytical separable model to the corresponding components of the separable approximation (16). Four examples are shown in Figs. 3 and 4. The model provides a good fit for the sample signal spectra in all frequency ranges. (Note that the ordinate scale is logarithmic, and therefore, the contribution to the mean squared error is small at high frequencies.) This finding is consistent with the applicability of the models of (6) and (7) in earlier studies of spatial and temporal statistics [2], [5], [7]–[9].

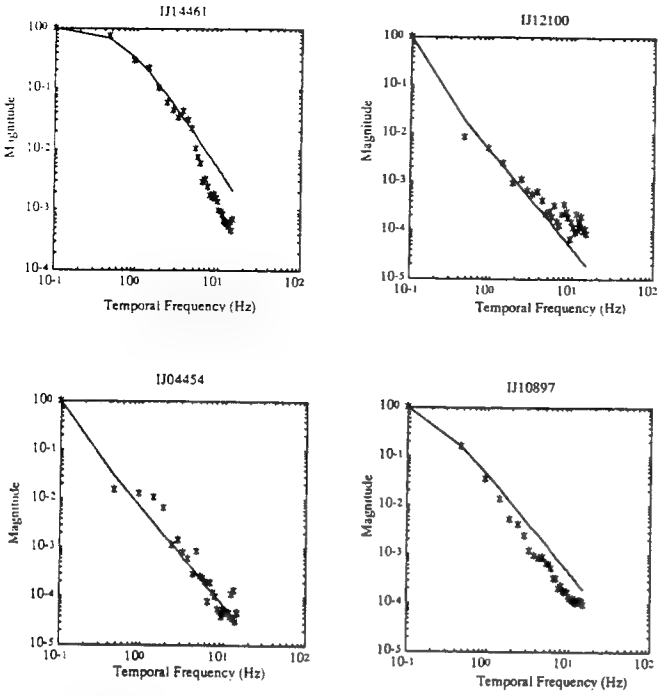


Fig. 4. Magnitude of the temporal component of the spectrum derived by the singular value decomposition (stars) compared with the analytical model (solid line). (Note different ordinate scales.)

VI. Discussion

We calculated the spatiotemporal power spectra of 14 image sequences to investigate whether these spectra are separable in space and time. Using a normalized index of separability, we show that a separable approximation for the spectra derived from the singular value decomposition extracts over 98% of the signal energy (Table I). We also investigated whether the space-time separable exponential model commonly used in the literature provides a reasonable description of the statistics of image sequences. This exponential model is equivalent to the space-time separable power spectrum model of (8). We show that this model provides a good analytical description of the spectrum of image sequences.

For this ensemble of image sequences, no sequence possessed both high spatial and high temporal frequencies (Fig. 2). This property may be a result of spatial blurring caused by motion. If so, it is not an inherent property of the image sequence but rather is caused by the low-pass temporal filtering of the camera. The visual system also temporally low-pass filters images (mainly due to photoreceptor integration time); therefore, this property holds true for a signal perceived by the visual system as well. This limitation on signal spatiotemporal bandwidth may be useful for perceptually based image coding and processing applications [14].

Applications of the model to image processing accrues both the advantages and limitations of using autocorrelation and power spectrum methods. As descriptions of images, the autocorrelation and power spectra are global in the sense that they represent a calculation averaged over the entire image or image sequence. This averaging does not retain the phase spectrum of images and removes local nonstationarities and, hence, specific local details of images. In addition, the separable model may not apply to local sections of image sequences even though the global spectrum of the sequence is separable. In those cases where the autocorrelation and power spectrum methods are applicable, the assumption of separability enables considerable mathematical simplicity. Any methods of image processing developed for spatial-only or temporal-only processing using (6) and (7) can be extended in a straightforward manner to spatiotemporal processing with (8).

References

[1] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[2] A. Rosenfeld and C. C. Kak, *Digital Picture Processing*. New York: Academic, 1982.

[3] M. V. Srinivisan, S. B. Laughlin, and A. Dubs, "Predictive coding: A fresh view of inhibition in the retina," *Proc. Roy. Soc. Lond. B*, vol. 216, pp. 427–459, 1982.

[4] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[5] D. J. Connor and J. O. Limb, "Properties of frame-difference signals generated by moving images," *IEEE Trans. Commun.*, vol. COM-10, pp. 1564–1575, 1974.

[6] L. E. Franks, "A model for the random video process," *Bell Sys. Tech. J.*, pp. 609–630, 1966.

[7] D. C. Coll and G. K. Choma "Image activity characteristics in broadcast television," *IEEE Trans. Commun.*, vol. COM-12, pp. 1201–1206, 1976.

[8] W. Chen and D. Hein "Recursive temporal filtering and frame rate reduction for image coding," *IEEE J. Spec. Areas Commun.*, vol. SAC-5, no. 7, pp. 1155–1165, 1987.

[9] A. K. Jain, "Partial differential equations and finite-difference methods in image processing, Part 1: Image representation," *J. Optim. Theory Appl.*, vol. 23, no. 1, pp. 65–91, 1977.

[10] S. Treitel and J. L. Shanks, "The design of multistage separable planar filters," *IEEE Trans. Geo. E.*, vol. GE-9, pp. 10–27, 1971.

[11] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing*. Englewood Cliffs, Prentice-Hall. 1984.

[12] R. C. Gonzalez and P. Wintz, *Digital Image Processing*. Reading, MA: Addison Wesley, 1987.

[13] J. E. Dennis and D. J. Woods, "Optimization in microcomputers: The Nelder-Meade simplex algorithm," in *New Computing Environments: Microcomputers in Large-Scale Computing* (A. Wouk, Ed.). New York: SIAM, 1987, pp. 116–122.

[14] A. B. Watson, "Perceptual-components architecture for digital video," *J. Opt. Soc. Amer. A*, vol. 7, no. 10, pp. 1943–1954, 1990.

Group 2: Signal propagation in the retina

Appendix D: Signal Sampling and Propagation through Multiple Cell Layers in the Retina: Modeling and Analysis Using Multirate Filtering, Journal of the Optical Society of America, Series A, Vol. 7, pp. 1463-1480, 1993

Appendix E: Conversion Between Parallel and Hierarchic Architecture Analysis Multirate Filter Banks, IEEE Transactions on Signal Processing, Vol. 40, pp. 2837-2841, 1992

Appendix F: Complexity and Filter Memory Requirements in Scaled Gaussian Hierarchic and Parallel Filter Banks, J. Visual Communications and Image Representation, Vol. 4, pp. 187-195 1993

Appendix D

Signal sampling and propagation through multiple cell layers in the retina: modeling and analysis with multirate filtering

Bennett Levitan and Gershon Buchsbaum

Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, 220 South 33rd Street, Philadelphia, Pennsylvania 19104-6392

Received May 27, 1992; revised manuscript received January 5, 1993; accepted January 11, 1993

The retina is a multilayered structure. Each layer consists of one or more classes of cell, each at its own density and with its own anatomic and physiologic properties. Signals converge from many cells in one layer onto single cells in another layer, and a signal from a single cell diverges to many cells in the next layer. In this methods paper we develop a general approach to retinal analysis and modeling that incorporates multiple cell classes, their densities, and related anatomic properties. The method is based on multirate filtering, a branch of signal processing in which signals of different sampling rates are manipulated. By drawing a correspondence between cell density and signal sampling rate, we define multirate models that incorporate different cell densities, convergence, divergence, variation in dendritic field shape, cell-to-cell variation in synaptic weights, and other anatomic features. We develop the multirate approach and apply it to the cat cone ⇒ cone bipolar CBb_1 ⇒ on-β ganglion cell pathway as an example. We calculate the spatial frequency responses of the CBb_1 and on-β cells based on the cone spatial frequency response and find that the attenuation of high frequencies in the cones prevents aliasing that would otherwise occur in CBb_1 and on-β cells. We compare the calculations with cat psychophysics. We show that the optics of the cat eye are insufficient in themselves for the prevention of aliasing in these cells; additional attenuation by the cone–cone gap junctions and the cone aperture is necessary. By including this postreceptoral filtering, we demonstrate that the highest spatial frequency that can be passed by the retina without aliasing is determined not always only by the densities of cones, bipolar cells, and ganglion cells but also by the synaptic and the dendritic weighting between these cells.

1. INTRODUCTION

The retina is a multilayered structure. Each layer consists of one or more classes of cell, each at its own density and with its own anatomic and physiologic properties. Signals converge from many cells in one layer onto single cells in another layer, and a signal from a single cell diverges to many cells in the next layer (Fig. 1). Recent studies of the cat retina have measured the detailed anatomic properties that are necessary to model this information flow for several classes of cells. Wässle *et al.* measured the convergences and the divergences between photoreceptors and type A and B horizontal cells.[1,2] Cohen and Sterling distinguished and modeled several classes of cone bipolar, their densities, and their convergences and divergences to cones and on-β ganglion cells.[3–6] Other applications that model retinal processing and coding and use actual convergences, divergences, and number of synapses can be found in Refs. 7–9. If explicit attention is paid to these and other properties, these models could accurately incorporate many anatomic details such as dendritic field shape and the number of synapses between cells. All these properties are highly dependent on the different densities of cell classes.

The purpose of this paper is to describe a method of retinal modeling that generalizes the multiple cell layer approach taken in these studies of the cat retina. Our motivation is to incorporate detailed anatomic properties into multilayered retinal models that can be easily analyzed. The method is based on a branch of signal processing known as multirate filtering. Multirate filtering concerns the manipulation, the filtering, and the analysis of signals in systems whose signals are not all at the same sampling rate. By drawing a correspondence between the density of an array of cells and the sampling rate of a signal, we can incorporate into retinal models different cell densities, convergence, divergence, variation in dendritic field shape, cell-to-cell variation in synaptic weights, and other anatomic features. Multirate models are easily manipulated analytically in both space and frequency domains, have computationally efficient implementations, and allow for closed-form solutions.[10–12] Multirate filtering permits an examination of the artifacts that result from a change in a signal's sampling rate. For this reason, it is particularly useful for the analysis of effects such as aliasing in a multilayered structure with different densities, such as the retina, and for the modeling of postreceptoral filtering. Some of the retinal properties that can be derived from this approach, such as the trade-offs in properties among parallel, hierarchic, and hybrid architectures, are reported in an early form in Refs. 13–15. Among the benefits of the multirate approach used here is that the models display some of the variation and the irregularity seen in retinal anatomy.

In Subsection 2.A of this paper we develop the method of modeling with multirate filtering. In Subsection 2.B we apply this technique to the cat cone ⇒ cone bipolar CBb_1 ⇒ on-β ganglion cell pathway and demonstrate some of the types of analysis permitted by multirate filtering. We show how spatial aliasing is prevented in this path-

0740-3232/93/071463-18$06.00

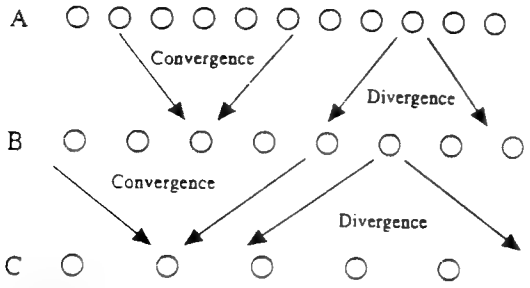


Fig. 1. Demonstration of multilayered structure of the retina. Rows indicate cells of classes A, B, and C, each of a different density. Circles represent the cells. The arrows indicate the range of cells that synapse between different layers. The average number of presynaptic cells that synapse on a postsynaptic cell is the convergence. The average number of postsynaptic cells on which a presynaptic cell synapses is the divergence.

way and compare our result with psychophysical measurements. We also use the model to estimate to what degree the optics of the cat eye alone are responsible for the prevention of aliasing in this pathway.

2. THEORY

A. Multirate Model

1. *One-Dimensional Case*

We draw a correspondence between anatomic and multirate signal processing parameters by equating the distance between signal samples to the average distance between cells. Table 1 lists the correspondences between anatomy and the model that we develop in this subsection and Subsections 2.A.2 and 2.A.3. Figure 2(a) illustrates the general synaptic arrangement based on which we develop the model. Consider cells of class A presynaptic to cells of class B. The densities of the two classes (in cells/unit length) are d_A and d_B, respectively. The average distance between cells is the nearest-neighbor distance (nnd). For a one-dimensional array of cells,

$$\mathrm{nnd}_A = 1/d_A. \tag{1}$$

In the multirate model we define the signals $x(n)$ and $y(m)$ with uniformly spaced samples so that they correspond to the one-dimensional arrays of A and B cells, respectively. The distance between samples in $x(n)$ is the sampling period (T_x). If we set the sampling periods as

$$T_x = \mathrm{nnd}_A, \qquad T_y = \mathrm{nnd}_B, \tag{2}$$

then the sampling rates of x and y equal the densities of the A and B cells, respectively. Each sample in $x(n)$ then corresponds to a different A cell, and each sample in $y(m)$ corresponds to a different B cell [Fig. 2(b)]. The image signals have different spatial variables because the n_0th sample in x does not correspond to the same location as that of the n_0th sample in y, except at the origin [Fig. 2(b)]. We address the issue of irregular sampling in the retina in Section 3.

To an excellent approximation, at a single adaptive state, many retinal cells sum their inputs linearly over both space and time.[16–18] For these cells the voltage produced in a B cell can be viewed as a weighted sum of the voltages in the presynaptic A cells, where the weights are determined by the number and the location of synapses, the details of the dendritic tree, the types of neurotransmitter, and the types of receptor. Since the cells sum linearly, one can simulate the operation of the synapses by convolving $x(n)$ with a filter of the appropriate weights. This approach is taken in many signal-processing-based models. However, since $x(n)$ and $y(m)$ are at different sampling rates, standard convolution is inadequate, as standard convolution assumes the same sampling rates in all the layers. Convolution between signals with different sampling rates requires multirate filtering. The use of multirate filtering allows for different sampling rates (densities) in different layers and hence arbitrary convergence/divergence ratios (see Subsection 2.A.1.b). We now summarize the relevant basics of multirate filtering and use them to develop the model.

The basic operations of multirate filtering are upsampling and downsampling.[10–12] Operating on a signal $x(n)$, an L-fold upsampler inserts $L-1$ zero-valued samples between adjacent samples in $x(\ )$ and decreases the sampling period of $x(\ )$ by a factor of L. Combined with the appropriate low-pass filtering, the upsampler increases the sampling rate of $x(\ )$ by a factor of L while maintaining the signal's form [Fig. 3(a)]. Operating on $x(n)$, an M-fold downsampler removes $M-1$ of every M samples in $x(\ )$ and increases the sampling period of $x(\ )$ by a factor of M. Combined with the appropriate low-pass filtering, a downsampler decreases the sampling rate of $x(\ )$ by a factor of M while maintaining the exact form of the signal [Fig. 3(b)]. Of importance is that, without the appropriate filtering, either resampler may induce aliasing (technically called imaging for the upsampler).[10–12] The insets in Fig. 3 demonstrate the anatomic situations that correspond to these resampling operations.

Technically, one can avoid the resampling operations in the model by setting both T_x and T_y to the greatest common divisor of nnd_A and nnd_B. This small sampling period ensures that both $x(\ )$ and $y(\ )$ have samples at the locations that correspond to every cell of classes A and B. However, this choice also results in artifactual filler

Table 1. Anatomic Parameters and Their Analogs in a Multirate Model

Anatomy	Multirate Model
Nearest-neighbor distance (nnd)	Sampling period T
Convergence C	Model convergence C_m
Divergence D	Model divergence D_m
Changes in cell densities	Downsampling/upsampling by M/L
Synaptic/dendritic weighting functions	Direct-form filter $h(u)$
Dendritic field radius	Circular filter radius R_{circ}
Dendritic fields and weights	Space-varying filters $g_m(n)$
Neurotransmitter/receptor gain	Filter gain K

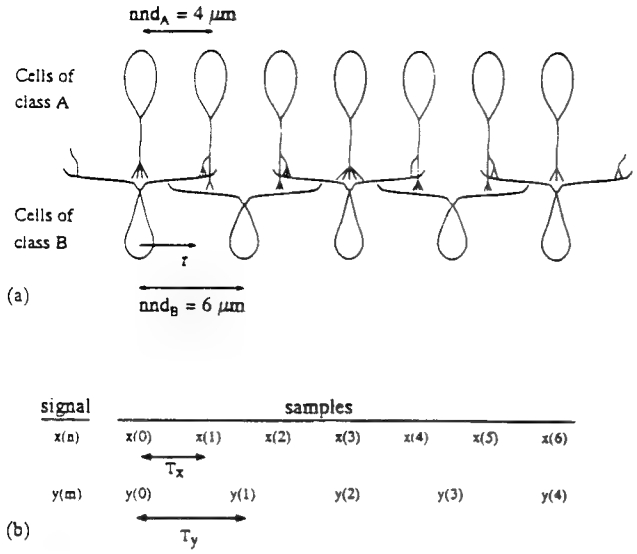


Fig. 2. Idealized description of synapses between two layers of cells and their depiction in the model. (a) Cartoon view of cells of class A presynaptic to cells of class B. Class A cells synapse on various numbers of B cells and provide different numbers of synapses to each. Offshoots on the terminal branches of an A cell indicate the number of synapses that it provides to the B cells beneath it. On the average, A and B cells are separated by $\mathrm{nnd_A}$ and $\mathrm{nnd_B}$, respectively. Here $\mathrm{nnd_A} = 4$ μm, and $\mathrm{nnd_B} = 6$ μm. (b) Signals $x(n)$ and $y(m)$ in the multirate model correspond to the arrays of A and B cells in (a). T_x and T_y are the sampling periods of $x(n)$ and $y(m)$, respectively.

samples in $x(\,)$ and $y(\,)$ that do not have corresponding cells in classes A and B. To prevent filler samples from interfering with the analysis and the interpretation of the model, we set the sampling periods of $x(\,)$ and $y(\,)$ so that the cells and the samples correspond.

To model the synaptic weighting between the cell arrays in Fig. 2(a), we use the system shown in Fig. 3(c), an L-fold upsampler followed by a filter and then by an M-fold downsampler. This sytem is a cascade of the systems in Figs. 3(a) and 3(b) with the filters combined. The use of both an upsampler and a downsampler permits the sampling rates of $x(n)$ and $y(m)$ to be related by any rational number and provides for any change in densities between class A and class B cells. In this case,

$$T_y = T_x M/L. \tag{3}$$

The resampling operations in Fig. 3 are the basic mathematical tools necessary for defining and manipulating multirate systems. There are several means by which one can implement these operations; for example, the three steps can be combined into a single-step operation that does not explicitly insert or remove samples.[10-12] We show in Subsection 2.A.2 that such a combined operation is analogous to the synaptic weighting in Fig. 2(a).

To derive M and L from anatomic parameters, we combine Eqs. (1)–(3) to obtain

$$\frac{M}{L} = \frac{T_y}{T_x} = \frac{\mathrm{nnd_B}}{\mathrm{nnd_A}} = \frac{d_A}{d_B}. \tag{4}$$

Equation (4) gives the ratio M/L in terms of the anatomic densities for the one-dimensional case. The actual values for M and L are given by the smallest pair of integers M and L that satisfy Eq. (4). There is often flexibility in the exact choice of M and L, since Eq. (4) can be exactly satisfied only when d_A/d_B is a rational number. Provided that M and L are relatively prime,[19] increasing M and L allows M/L to approximate a given d_A/d_B better and causes the sampling rates T_x and T_y to match $\mathrm{nnd_A}$ and $\mathrm{nnd_B}$ more closely. We discuss other effects of an increase in M and L when we consider the two-dimensional case in Subsection 2.A.3.

In Fig. 3(c) $h(u)$ is a discrete spatial filter with N_h samples that emulates the synaptic and dendritic weighting. In deriving $h(u)$ from anatomic parameters, one must consider three aspects of the filter: (a) functional form, (b) spatial extent, and (c) gain. We now discuss the relevant anatomy and its correlate in the model for each aspect.

a. Functional Form In Fig. 2(a) the number of offshoots on the terminal branches of an A cell indicates the number of synapses that it provides to the B cells beneath it. The number of these synapses is characterized by the synaptic weighting function (swf).[8] Calculated from retinal data, $\mathrm{swf}(x, y)$ gives the average number of synapses between a B cell and an A cell that synapses at location (x, y) relative to the center of the B cell's dendritic field. Figure 4(a) shows the swf for the one-dimensional array of cells in Fig. 2(a). In this example $\mathrm{swf}(r) = 4 - 0.5|r|$, where r is in micrometers.

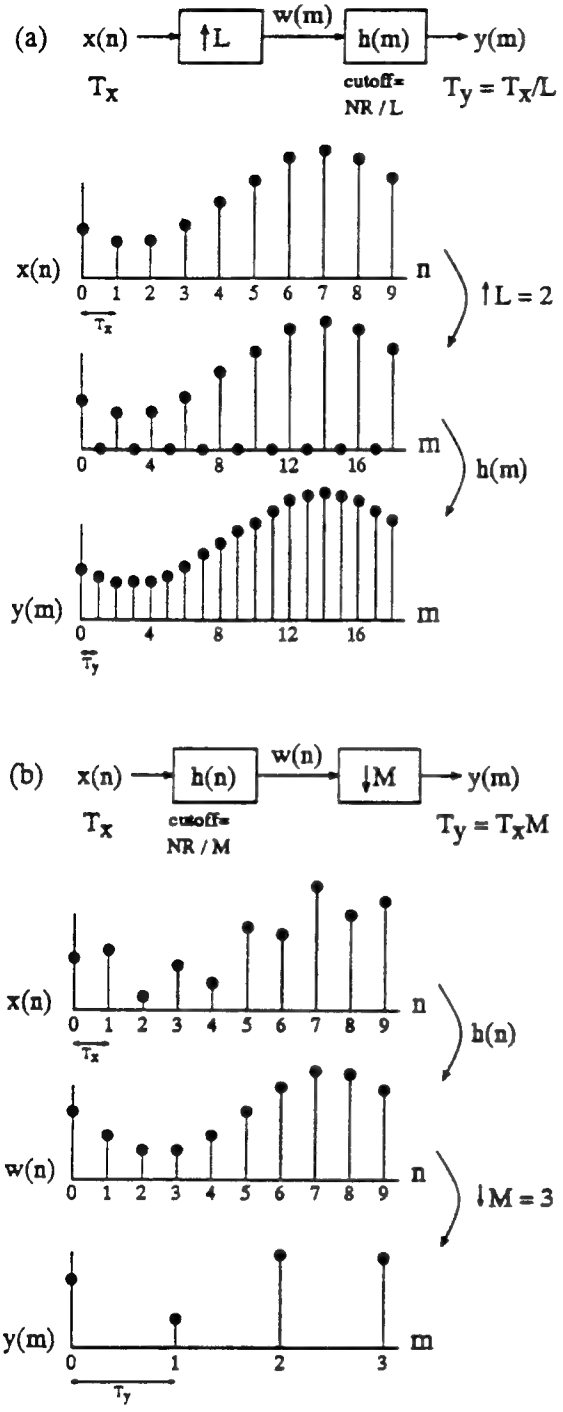
The change in voltage induced in a B cell by the A cells presynaptic to it is characterized by the dendritic weighting function (dwf). $\mathrm{dwf}(x, y)$ gives the average change in voltage in the soma of a B cell induced by a unit injection of current from a presynaptic A cell that synapses at location (x, y) relative to the center of the B cell's dendritic field. The dwf reflects branching in the dendrites, dendritic diameters, and membrane and cytoplasmic characteristics.[20] Though the swf and the dwf may be of any form, in practice they are often assumed to be on average circularly symmetric, such that $\mathrm{swf}(x, y)$ and $\mathrm{dwf}(x, y)$ are functions of the radius $(x^2 + y^2)^{1/2}$. With this assumption the weight between a presynaptic and a postsynaptic cell is a function of the distance between them.

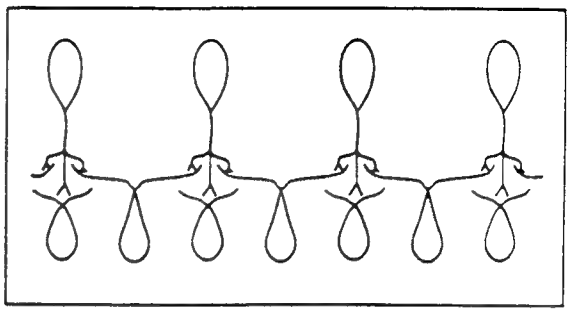
Assuming that all synapses of the same type contribute equally to the postsynaptic potential, the functional form of $h(u)$ is proportional to both the swf and the dwf:

$$h(u) = K\,\mathrm{dwf}(uT_x/L)\mathrm{swf}(uT_x/L), \tag{5}$$

where uT_x/L is the continuous distance that corresponds to discrete distance u at the filter's sampling rate. If we set $\mathrm{dwf}(r) = 1$, the functional form of the filter that corresponds to the swf in Fig. 4(a) is given by $h(u) = 4 - |u$ [Fig. 4(b)].

By means of several mechanisms retinal cells adapt to the signals that they carry. They change their overall gains, alter the shape of their receptive fields, and adjust their chromatic and temporal properties according to present and past input.[16,18,21,22] The spatial components of these changes manifest themselves in the dwf's and the overall gains of the cells. In contrast, swf's are relatively constant. Our development below is for one level of adaptation, so the gains and the dwf's remain constant. One can study different adaptive states by changing these components to match the different states or by using adaptive filters or nonlinearities.[23-26]





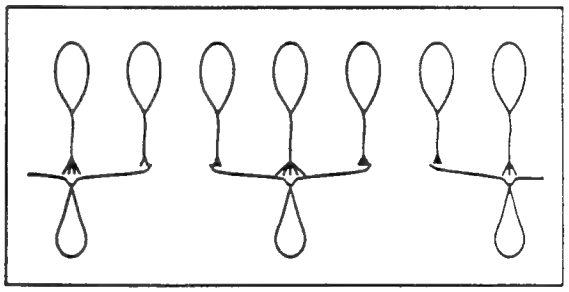


Fig. 3 Continues on facing page.

b. Spatial Extent The number of class A cells that synapse on a class B cell varies from B cell to B cell. Similarly, the number of B cells to which an A cell is presynaptic varies from A cell to A cell. The average number of A cells that synapse on a single B cell is the convergence (C) (Fig. 1).[27] In the early layers of vision, convergence reflects the degree to which a cell gathers information over the visual field. The average number of B cells on which a single A cell synapses is the divergence (D) (Fig. 1).[27] Divergence is one of the mechanisms by which a point of visual information spreads laterally among the cells. In Fig. 2(a) $C = 2.5$ and $D = 1.667$. Clearly, convergence and divergence are related. Freed *et al.*[27] have shown that, for two arrays of cells as in Fig. 2(a),

$$C/D = d_A/d_B . \tag{6}$$

The filter's spatial extent N_h is set so that the convergence and the divergence in the model match those of the anatomy. After the upsampler in Fig. 3(c) $L - 1$ of every L samples in $v(u)$ are zero. In the model the convergence is the average number of samples from $x(n)$ that $h(u)$

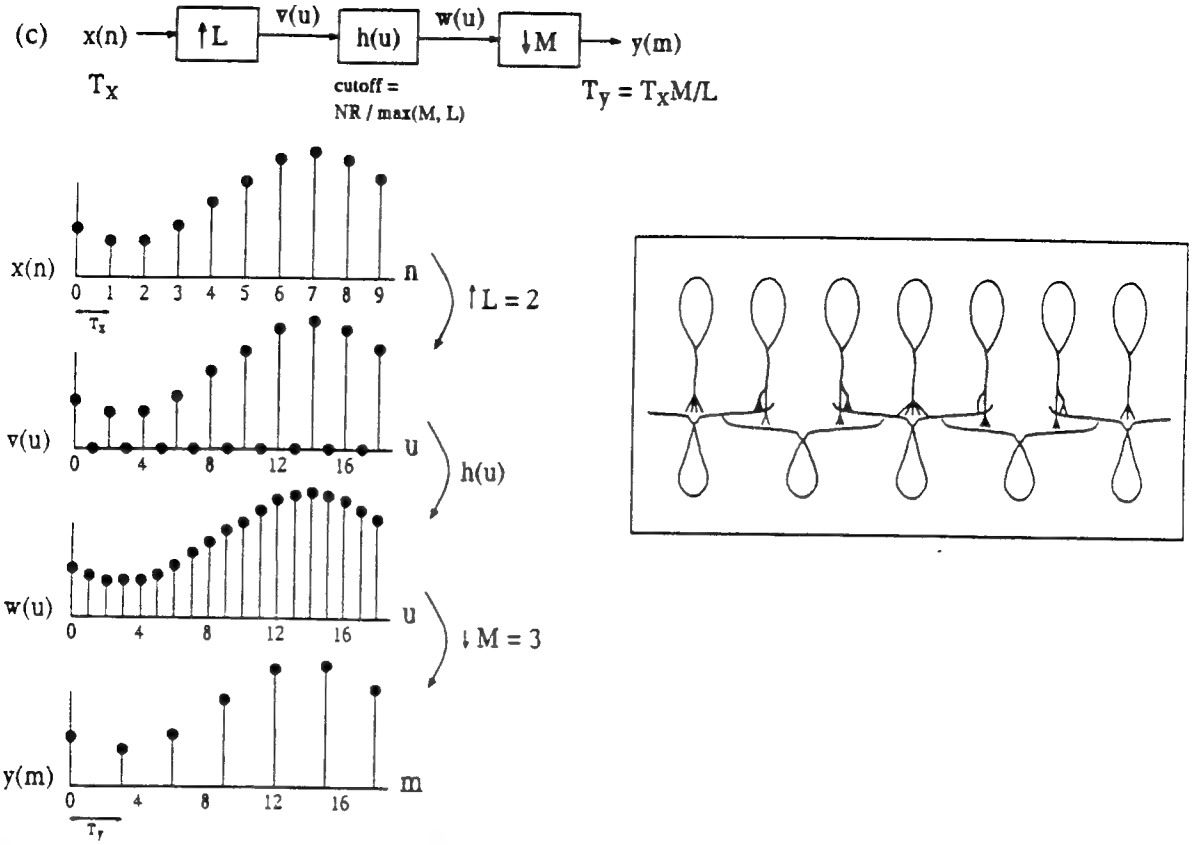


Fig. 3. Resampling operations with appropriate low-pass filtering. NR stands for Nyquist rate. (a) The L-fold upsampler, indicated by the box with the upward-pointing arrow, inserts $L - 1$ zero-valued samples between each adjacent samples in $x(n)$. In the example shown here, $L = 2$. By causing the signal to drop to and from zero so often, the upsampler introduces high-frequency components. Placing a low-pass filter with a cutoff at NR/L after the upsampler removes these components. As is shown in the inset, this operation is analogous to the effect of synapses between two layers of cells, where the second layer's density is L times the density of the first. (b) The M-fold downsampler, indicated by the box with the downward-pointing arrow, removes $M - 1$ of every M samples in $w(n)$. In the example shown here, $M = 3$. By bringing spatially distant samples together, the downsampler moves frequency components in $w(n)$ to higher frequencies. Components that are moved beyond the Nyquist rate alias. Placing a low-pass filter with a cutoff at NR/M before the downsampler prevents the aliasing. As is shown in the inset, this operation is analogous to the effect of synapses between two layers of cells, where the second layer has $1/M$ the density of the first. (c) L-fold upsampling and M-fold downsampling combine to change the sampling period by a factor of M/L. Placing a low-pass filter with a cutoff at NR/max(M, L) between the resamplers prevents the aliasing. As is shown in the inset, this operation is analogous to the effect of synapses between two layers of cells, where the second layer has L/M times the density of the first. In the example shown here, $L = 2$ and $M = 3$, which make the operation correspond to the synapses and the signals in Fig. 2.

touches or the average number of nonzero samples in $v(u)$ that $h(u)$ touches. Thus a filter length of N_h yields

$$C_m = N_h/L, \tag{7}$$

where C_m is the convergence in the model. The divergence in the model, D_m, is the average number of samples in $y(m)$ that are calculated at least in part from one sample of $x(n)$ and is given by

$$D_m = N_h/M. \tag{8}$$

If anatomic measurements satisfy Eq. (6) exactly and if M/L satisfies Eq. (4) exactly, then $N_h = CL$ yields $C_m = C$ and $D_m = D$. This case holds for the example in Fig. 2(a), where setting $N_h = 5$ yields $C_m = 2.5$ and $D_m = 1.667$. When Eqs. (4) and (6) are not satisfied exactly, as is typically the case given the precision of anatomic measurements, no value of N_h yields both $C_m = C$ and $D_m = D$. As is shown in Subsection 2.B below, varying N_h results in trade-offs between the model's matching the anatomic convergence more closely and the model's matching the anatomic divergence more closely.

c. Gain The gain of $h(u)$ is represented by the coefficient K in Eq. (5). The gain depends on the types of neurotransmitter and receptor in the synapse. When only one pair of cells is being modeled, K can generally be ignored except for its sign—excitatory synapses require positive K, and inhibitory synapses require negative K. When a model incorporates several sets of cells in parallel, the relative gains of the filters must be included. If it is intended that the values of $x(n)$ and $y(m)$ match the voltages in A and B cells, respectively, then we can set K by comparing the presynaptic and postsynaptic voltages for a spatially constant input to the A cells, using

$$K = \frac{L^2 \times (\text{voltage in B cells})/(\text{voltage in A cells})}{\sum_u [\text{dwf}(uT_x/L)\text{swf}(uT_x/L)]}, \tag{9}$$

where the summation normalizes for the dc component of the convolved dwf and swf. If we intend that the values of $x(n)$ and $y(m)$ relate only linearly to voltages in A and B cells, respectively, then Eq. (9) can be used or the K's for the parallel filters can be made proportional to the total

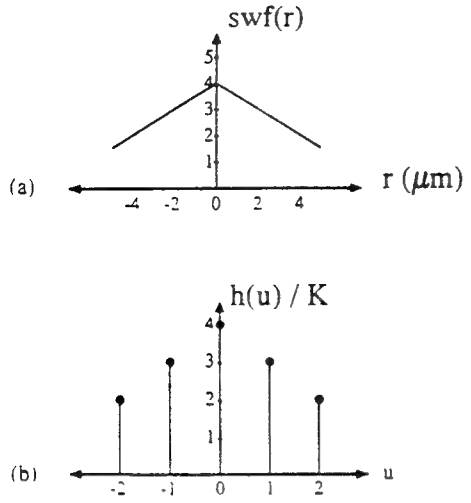


Fig. 4. Plot of the synaptic weighting function (swf) for the synapses in Fig. 2(a) and its multirate model analog: (a) swf, (b) corresponding direct-form filter $h(u)$.

number of synapses on the postsynaptic cells. For the swf in Fig. 4(b) K is set to unity.

2. Space-Varying Filter Implementation

The system in Fig. 3(c), in which the upsampler, the filter, and the downsampler operate separately, is known as the direct-form filter implementation. This implementation is a necessary and a standard tool for defining and analyzing multirate models.[11] However, there are obviously no explicit resamplers in the retina. The multirate filter implementation that corresponds to the anatomy is the space-varying filter implementation.[10-12] It performs the same operations as those of the direct-form implementation, but it combines the resampling and the filtering into one step. Also known as the time-varying or the polyphase filter implementation, the space-varying filter implementation is so named because it uses different filters along the spatial axis. Several filters are necessary because the sampling rates of $x(n)$ and $y(m)$ differ. Samples of these signals are not always aligned [Fig. 2(b)], and different filters must be used in different locations for the direct computation of $y(m)$ from $x(n)$. Figure 5 demonstrates the space-varying implementation for the example used in Figs. 2, 3(c), and 4. We compute $y(m)$ directly from $x(n)$. The implementation cycles between a three-weight filter $g_0(n)$ (which computes odd-numbered outputs) and a two-weight filter $g_1(n)$ (which computes even-numbered outputs). In the general one-dimensional case the space-varying implementation cycles between L different $g_m(n)$, which are given by

$$g_m(n) = h(nL + mM \bmod L), \tag{10}$$

where $a \bmod b$ is the remainder of a divided by b. The $g_m(n)$ vary in length and are composed of the samples of $h(u)$ taken in a specific order.[11,12]

Though it is designed with average measures of anatomy (convergence, divergence, swf, and dwf), the space-varying implementation of the multirate model demonstrates variation reminiscent to that in the retina. Comparing the model in Fig. 5(a) with the simplified anatomy in Fig. 2(a), we see that the space-varying filters $g_m(n)$ are analogous to the dendritic fields of varying sizes. The samples in $x(n)$ that compute a given sample in $y(m)$ corresponds to the class A cells that synapse on the corresponding class B cell. Not every B cell has the same convergence, just as not every filter $g_m(n)$ has the same size. The anatomic convergence C is the average convergence for all the B cells, and the model convergence C_m is the average size of all the filters $g_m(n)$. Different cells receive different number of synapses, and different space-varying filters have different weight values. We discuss this multirate variation more fully in Subsection 2.A.3. Table 1 summarizes the correspondence between anatomic properties and parameters in the model.

3. Two-Dimensional Case

For a two-dimensional array of cells, d_A is measured in cells/unit area, and the nnd is defined differently depending on whether we consider the cells to be packed in a rectangular, a hexagonal, or other sampling array. Our development is for a rectangular array, though hexagonal and other sampling schemes can be incorporated into multirate filtering.[28,29] For rectangularly packed A cells the nnd is defined as

$$\mathrm{nnd}_A = 1/\sqrt{d_A}. \tag{11}$$

We can derive the two-dimensional model by performing the multirate operations separately along each axis. The input signal $x(n_1, n_2)$ is upsampled by L along both axes, filtered by direct-form filter $h(u_1, u_2)$, and downsampled by M along both axes, which forms output $y(m_1, m_2)$. As in Subsection 2.A.1, $T_y = T_x M/L$. We find M/L by setting the sampling rates T_x and T_y equal to the nnd's as in Eq. (2). Combining Eqs. (2), (3), (6), and (11) yields

$$\frac{M}{L} = \frac{T_y}{T_x} = \frac{\mathrm{nnd}_B}{\mathrm{nnd}_A} = \left(\frac{d_A}{d_B}\right)^{1/2} = \left(\frac{C}{D}\right)^{1/2}. \tag{12}$$

Equation (12) differs from the corresponding one-dimensional relations [Eqs. (4) and (6)] in that the densities, the convergence, and the divergence in Eq. (12) are for a two-dimensional array of cells. As in Subsection 2.A.1, the

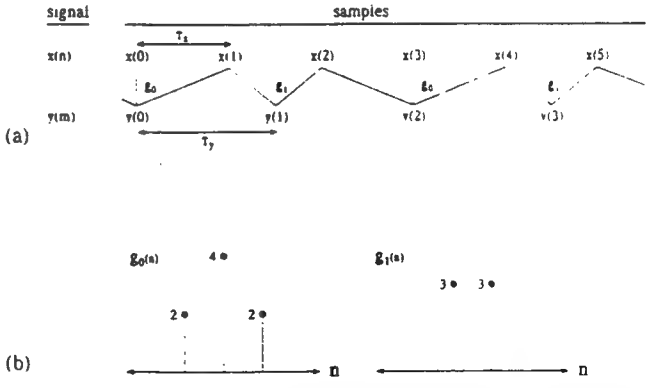


Fig. 5. Space-varying filter implementation of Fig. 3(c) for $L = 2$, $M = 3$, and filter $h(u)$ with five weights. The model shown here corresponds to the synapses and the signals in Fig. 2. (a) The samples are from input $x(n)$ and output $y(m)$. The lines show which samples from $x(n)$ are multiplied by filter weights for calculating samples in $y(m)$. A comparison with Fig. 2(a) shows that the samples in $x(n)$ that compute a given sample in $y(m)$ correspond to the class A cells that synapse on the corresponding class B cell. (b) Space-varying filters $g_0(n)$ and $g_1(n)$ for (a), which correspond to the swf and the direct-form filter in Fig. 4.

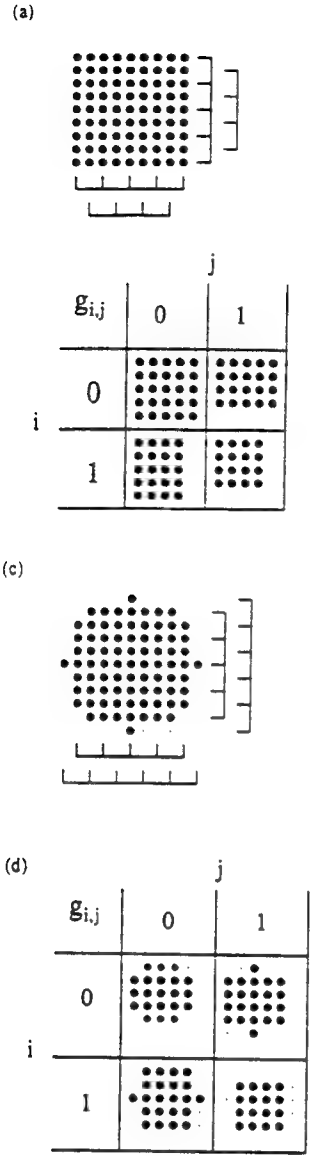


Fig. 6. Examples of square direct-form and circular direct-form filters and their corresponding resulting space-varying filters for $L = 2$ and $C = 20.25$. The large dots indicate filter weights of any value. (a) The square direct-form filter is 9×9 weight. The graticules show which samples are selected along each axis from the direct-form filter for constructing the space-varying filters. (b) The space-varying filters that correspond to the square filter are 5×5, 5×4, 4×5, and 4×4 weights. (c) Circular direct-form filter. The small dots are zero-valued weights. (d) Space-varying filters that correspond to the circular filter.

actual values for M and L are given by the smallest pair of integers that satisfy Eq. (12).

The functional form of the filter is proportional to both the swf and the dwf. Extending Eq. (5) to the two-dimensional case leads to

$$h(u_1, u_2) = K\,\mathrm{dwf}\left(\frac{u_1 T_x}{L}, \frac{u_2 T_x}{L}\right)\mathrm{swf}\left(\frac{u_1 T_x}{L}, \frac{u_2 T_x}{L}\right). \quad (13)$$

If $h(u_1, u_2)$ consists of $N_h \times N_h$ samples, then the model's convergence and divergence are given by

$$C_m = N_h^2/L^2, \quad (14)$$

$$D_m = N_h^2/M^2, \quad (15)$$

respectively. We set N_h to give the best match between the model's convergence and divergence and those of the anatomy. For two-dimensions there are L^2 space-varying filters given by

$$g_{m_1,m_2}(n_1, n_2) = h(n_1 L + m_1 M \bmod L, n_2 L + m_2 M \bmod L), \quad (16)$$

which is the two-dimensional equivalent of Eq. (10). Each $g_{m_1,m_2}(n_1, n_2)$ is composed of a different selection of samples from $h(u_1, u_2)$. The filters vary in length along each axis, are square or rectangular in shape, and have an average of C_m samples. The number of samples along each axis is either the smallest integer greater than N_h/L or the largest integer less than N_h/L. For example, the four space-varying filters that result from a 9×9-weight $h(u_1, u_2)$ for $L = 2$ and any M are 5×5, 5×4, 4×5, and 4×4 weights [Figs. 6(a) and 6(b)]. Their average number of weights is 20.5, which matches the convergence C_m, as one would expect.

To obtain space-varying filters that match the shapes of dendritic fields more closely [Fig. 10(b) below], we can make the direct-form filter circular in shape [Fig. 6(c)]. This approach gives space-varying filters that are on average circular [Fig. 6(d)] instead of rectangular [Fig. 6(b)]. The circular filter boundary is defined by

$$h(u_1, u_2) = 0 \quad \text{for} \quad (u_1^2 + u_2^2)^{1/2} > R_{\mathrm{circ}}, \quad (17)$$

where we set the radius R_{circ} for the best match between C_m and C and between D_m and D. For circular filters the model convergence and divergence are given by

$$C_m = (\text{number of nonzero weights in } h)/L^2, \quad (18)$$

$$D_m = (\text{number of nonzero weights in } h)/M^2, \quad (19)$$

respectively. The two direct-form filters in Fig. 6 have the same convergence, but the circular filter is slightly larger in compensation for the zero-valued weights.

When $(d_A/d_B)^{1/2}$ is irrational, we can set M and L to satisfy Eq. (12) with an arbitrary degree of accuracy. The consequences of a larger L are a larger filter size $N_h \times N_h$ and a greater variety of space-varying filters g_{m_1,m_2}. If $h(u_1, u_2)$ is square, the g_{m_1,m_2} will all differ in their weights but will never have more than four different sizes and shapes. If $h(u_1, u_2)$ is circular, the g_{m_1,m_2} will be of a variety of shapes and sizes depending on locations of zero-valued weights in $h(u_1, u_2)$. Thus changing M and L while keeping M/L approximately constant provides a means to improve the matches between the ranges of convergence and divergence in the model and those in anatomy.

Other sources of flexibility in the matching of variation in the model with that in anatomy arise when one ensures that the space-varying filters completely tile the presynaptic layer. Tiling refers to the degree to which filters touch every sample in their two-dimensional input array. The L^2 filters, constrained to match average anatomic properties, in some cases need additional variation to touch every input sample. Incomplete tiling corresponds to a situation in which presynaptic cells are not synapsed on by any postsynaptic cells. We demonstrate incomplete tiling in Subsection 2.B. There are at least three methods that introduce the variation needed for complete tiling:

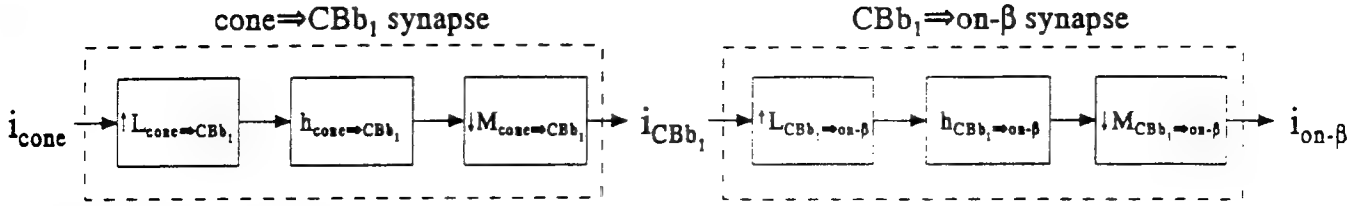


Fig. 7. Block diagram of the cone ⇒ cone bipolar CBb_1 ⇒ on-β ganglion cell pathway in the multirate model; i_{cone}, i_{CBb_1}, and $i_{on\text{-}\beta}$ refer to the images in the cone, CBb_1, and on-β cell arrays, respectively.

Table 2. Densities and Nearest-Neighbor Distances of Cones, CBb_1 Cells, and on-β Cells at 1° Eccentricity[a]

Cell Class	Density (mm^{-2})	Nearest-Neighbor Distance (μm)
Cones	24,200	5.2
CBb_1	6,100	10.2
On-β	1,860	25 (est.)

[a]Refs. 4 and 30.

(1) Vary the M/L ratio within the anatomic range for measurements of $(d_A/d_B)^{1/2}$. If we change L and hence the number and the variety of space-varying filters, tiling may become more nearly complete. While it is easy to implement, because of the limited range of $(d_A/d_B)^{1/2}$ this method may not work in all cases.

(2) Permit the direct-form filter to cycle among several asymmetric weighting functions that are on average symmetric. The average symmetry ensures that the filters act isotropically. Methods (1) and (2) increase the range of nonzero weights (convergences) of the space-varying filters and can make the filters better mimic the variability of retinal dendritic fields. Method (2) requires performing a different analysis for each direct-form filter used.

(3) Use different patterns of sampling in different layers, such as rectangular and hexagonal.

We discuss examples of all these methods in Subsection 2.B.

B. Application: Multirate Model of Cat Cone ⇒ Cone Bipolar b_1 ⇒ on-β Ganglion Cell Pathway

1. *Deriving the Model from Anatomy*

The cat cone ⇒ cone bipolar ⇒ on-β ganglion cell pathway is an ideal system for examination with multirate filtering, since much of its anatomy is known in great detail. The cone bipolars are of five morphological types, CBb_1–CBb_5, of which CBb_1 contributes more than half of the cone bipolar input to the on-β cell.[4] Here we model the pathway involving CBb_1 at 1° eccentricity. The two-stage model is shown in Fig. 7. The model's input is the output of the cone array, which in our analysis is either the signal formed by cones, horizontal cells, and the eye's optics or an arbitrary test signal. Other cells such as CBb_2–CBb_4 and amacrines also transmit signals from cones to on-β cells. These signals add to or may alter the CBb_1 signal after it has reached the on-β cells, but there is no difficulty in analyzing the CBb_1 path separately. Horizontal cells are partly responsible for generating the cone signal but do not have any influence on the signal once it has entered the bipolar cells. For clarity we will refer to the resampling factors in the model as M and L without subscripts, using the context to identify the particular synapse.

The densities for these cells are listed in Table 2. For the two synaptic steps involved, Table 3 lists the anatomic convergences, divergences, C/D ratios, and density ratios. Since the C/D ratios are not especially close to the density ratios, we use the more reliable density measurements to set M/L, then find the circular filter sizes that best match the model to the anatomic convergences and divergences.

a. Cone ⇒ CBb_1 Synapse For the cone ⇒ CBb_1 synapse (left-hand side of Fig. 7), Eq. (12) gives $M/L = \sqrt{3.97}$, for which $M = 2$ and $L = 1$ is a close match. Since $L = 1$, there is only one space-varying filter, which is identical to the direct-form filter. Setting $R_{circ} = 1$ results in a filter with five nonzero samples in a plus-shaped (+) pattern. Equations (18) and (19) yield $C_m = 5$ and $D_m = 1.25$, which give a close match between model and anatomic convergences and divergences. For this synapse observation of the anatomic convergence indicates that it ranges between 4 and 7, with an average of 5.1.[3] Because there is only one space-varying filter, convergence in the model does not vary from 5.0. The anatomy has more variation than does the model for this synapse, though the choice of other values for M and L could provide variation in the model's convergence while still matching the C/D ratio.

This synapse also provides an example of the tiling issue discussed in Subsection 2.A.3. As Fig. 8(a) shows, a plus-shaped (+) symmetric filter with five weights and the ratio $M/L = 2/1$ does not completely tile the cone layer. One of four cone samples is missed. We do not demonstrate it here, but varying M and L does not complete tiling for this synapse. Values of M/L close to $\sqrt{3.97}$ [Eq. (12)] yield tilings that are better than those provided by $M/L = 2/1$, but they are still incomplete. One solution that results in complete tiling is to cycle between two direct-form

Table 3. Anatomic Properties for the Cone ⇒ CBb_1 and CBb_1 ⇒ on-β Synapses at 1° Eccentricity[a]

Synapse	Convergence	Divergence	Convergence/Divergence	Density Ratio	Convergence Range[b]
Cone ⇒ CBb_1	5.1	1.2	4.25	3.97	4–7
CBb_1 ⇒ on-β	6–7	3	2–2.33	3.28	6–7

[a]Refs. 3, 5, and 6.
[b]The convergence range refers to the observed minimum and maximum number of presynaptic cells that converge on a postsynaptic cell for that synapse

filters, one plus shaped and one cross shaped (×) [Fig. 8(b)]. These filters are rotated and radially scaled versions of each other. Since $L = 1$, the space-varying filters are identical to the direct-form filters. Figure 8(c) shows a solution that introduces more variability in filter shape. Here we cycle among four different filters that are on average symmetric. Many other combinations of filters are possible. Another solution is to change the sampling in the on-β array to a slightly compressed hexagonal grid [Fig. 8(d)].[31] In this case a single plus-shaped filter touches every cone. These solutions show that varying from a single direct-form filter based on anatomic averages is sometimes needed for complete tiling of a presynaptic layer. In the frequency-domain analysis in Subsection 2.B.2 we use the tiling solution shown in Fig. 8(b).

Neither the swf nor the dwf is known for the cone ⇒ CBb_1 synapse. Since the CBb_1 dendritic field is so narrow,[5] we assume that both functions are constant. From Eqs. (13) and (17) the plus-shaped and cross-shaped direct-form filters are

$$h^{+}_{\text{cone}\Rightarrow\text{CBb}_1}(u_1, u_2) = \begin{cases} K & u_1^2 + u_2^2 \le 1 \\ 0 & \text{otherwise} \end{cases},$$

$$h^{\times}_{\text{cone}\Rightarrow\text{CBb}_1}(u_1, u_2) = \begin{cases} K & |u_1| = 1 \text{ and } |u_2| = 1 \\ K & u_1 = u_2 = 0 \\ 0 & \text{otherwise} \end{cases}, \quad (20)$$

respectively. These filters are shown in the insets of Fig. 9(a).

b. CBb_1 ⇒ on-β Synapse For the CBb_1 ⇒ on-β synapse (right-hand side of Fig. 7), Eq. (12) gives $M/L = \sqrt{3.28}$, for which $M = 16$ and $L = 9$ match well. Because of the large difference in the C/D and density ratios, no value of R_{circ} permits an exact numerical match of the model to both the anatomic convergence and the anatomic divergence. A value of R_{circ} that gives a C_m within the range of C also gives a D_m that is significantly less than D. While an increase in R_{circ} will increase the model's divergence, it will also bring the model's convergence out of the anatomic range 6–7. We set $R_{\text{circ}} = 13.35$, which yields a direct-form filter with 561 nonzero weights, a C_m of 6.93, and a D_m of 2.19. These values also permit complete tiling of the CBb_1 array by the space-varying filters. While the observed convergence for this synapse ranges from 6 to 7, the space-varying filters have between four and nine weights. For this synapse the model has more variation than does the anatomy. However, since the observed range is based on only three on-β cells, it is likely that the actual anatomic convergence varies more widely than the range 6–7. Because the M/L ratios do not exactly match the density ratios for either synapse, the model sampling periods are not identical to the anatomic nnd's. We set $T_{\text{cone}} = \text{nnd}_{\text{cone}} = 6.43$ μm. By Eq. (3), $T_{\text{CBb}_1} = 12.8$ μm and $T_{\text{on-}\beta} = 23.2$ μm. These values and the corresponding model densities are listed in Table 4.

Smith and Sterling argued that electrotonic decay along on-β cell dendrites at 1° eccentricity is insignificant, and therefore the dwf for the CBb_1 ⇒ on-β synapse is constant.[8] The swf is not known in detail, but anatomic measurements show that CBb_1 cells near the middle of the on-β dendritic field tend to give many contacts (12–33) and that cells near the edge of the field tend to give few contacts (3–4).[3] The radius of the on-β cell's dendritic field is approximately 20 μm.[3] Since, for this synapse, the variation in number of synaptic contacts is represented

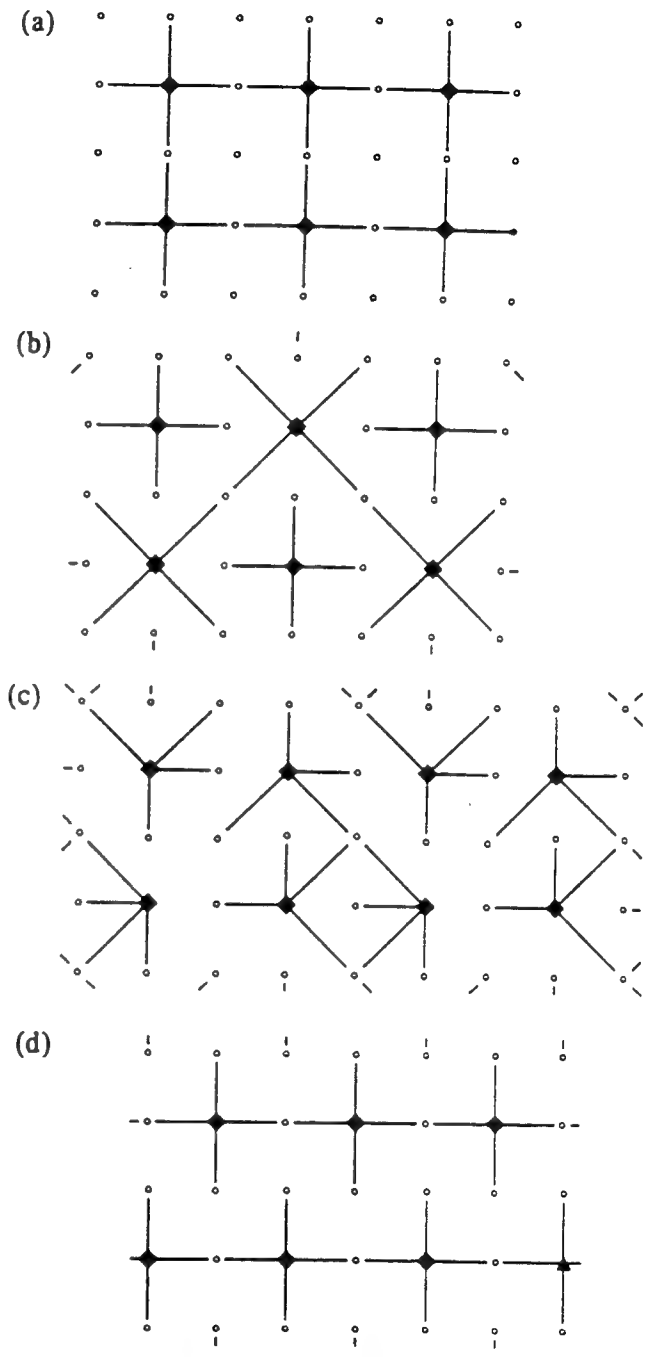
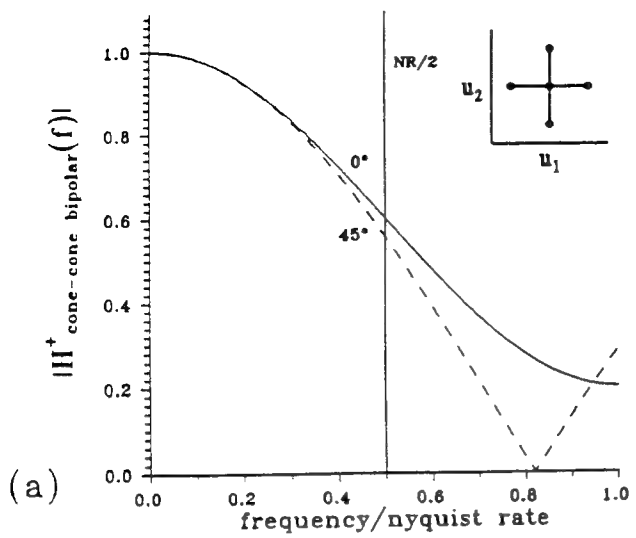
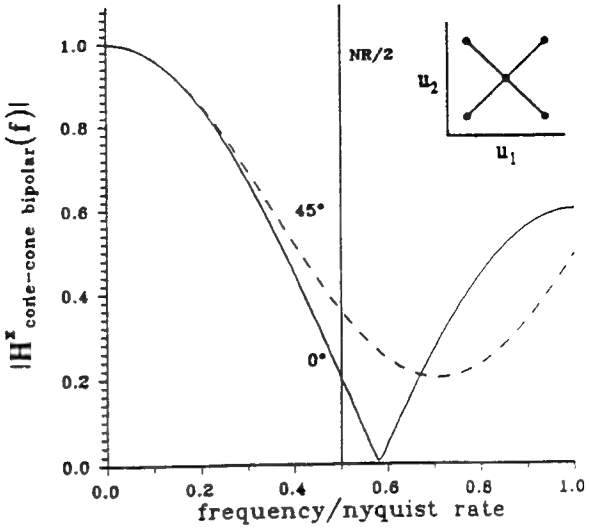
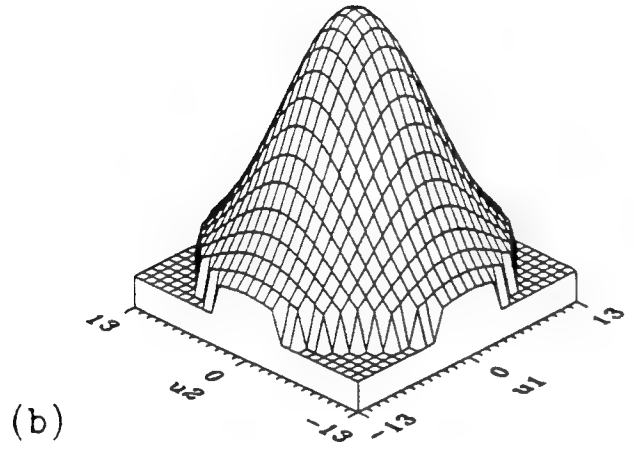
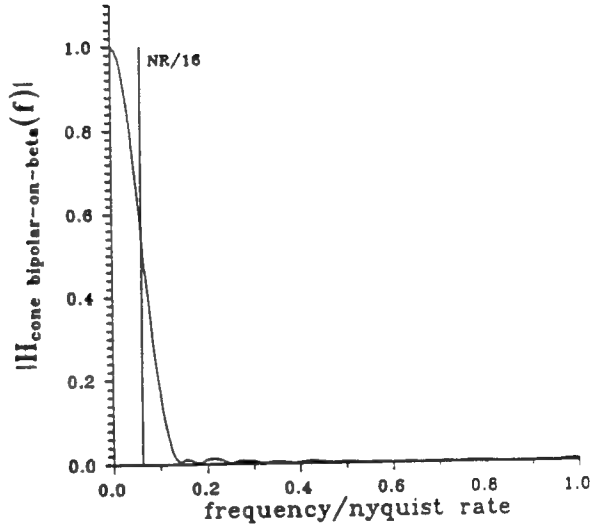


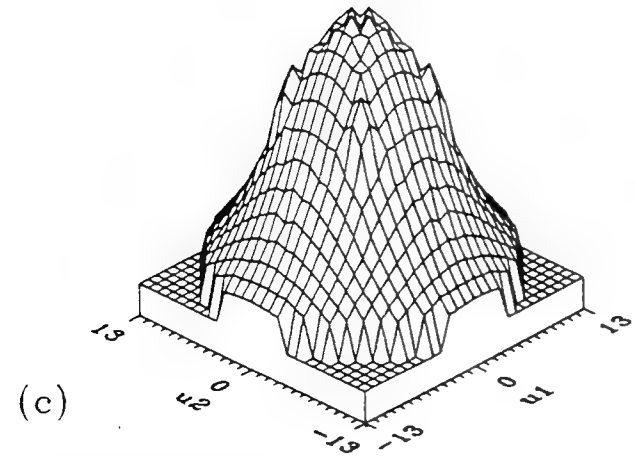
Fig. 8. Demonstration of tiling issue and solutions for the cone ⇒ CBb_1 filter. Filled diamonds indicate the CBb_1 sample, and small open circles indicate the cone samples touched by the CBb_1 samples. The contacts are shown explicitly by the lines. The resampling operations give the CBb_1 array one quarter of the density of the cone array. (a) Cone and CBb_1 samples and filter as defined in the text. The five-weight, plus-shaped filters touch only three of every four cone samples. (b) Cycling among several direct-form filters that are on average symmetric completely tiles the cone array. Here two filters are used. (c) Same method as in (b) but here the cycling is among four different filters. (d) Changing from rectangular sampling to another sampling scheme (in this case, compressed hexagonal sampling for the CBb_1 array) permits tiling of the cone array by a single, symmetric filter.











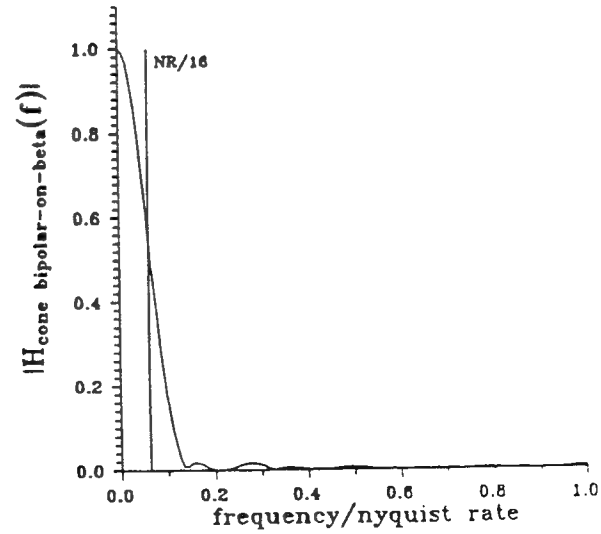


Fig. 9. Direct-form filters for the model. Spatial weighting functions are shown in the insets of (a) and on the left-hand sides of (b) and (c), and cross sections of their DTFT's are shown in (a) and on the right-hand sides of (b) and (c). The frequency axes are in units of normalized radian frequency and extend from zero to the Nyquist rate for each filter. The vertical lines show the cutoffs required for preventing aliasing. (a) $h^{+}_{\text{cone}\Rightarrow\text{CBb}_1}$ and $h^{\times}_{\text{cone}\Rightarrow\text{CBb}_1}$ filters, DTFT's are shown along a 0° line (ω_x axis) and along a 45° line ($\omega_x = \omega_y$); (b) $h_{\text{CBb}_1\Rightarrow\text{on-}\beta}(u_1, u_2)$, DTFT along a 0° line; (c) rescaled $h_{\text{CBb}_1\Rightarrow\text{on-}\beta}(u_1, u_2)$, DTFT along a 0° line.

Table 4. Densities, Sampling Periods, and Nyquist Rates for Cell Arrays in the Multirate Model

Cell Class	Density in Model (mm^{-2})	Sampling Period in Model (μm)	Nyquist Rates in Model (cycles/degree)
Cones	24,200	6.43	17.1
CBb_1	6,050	12.86	8.56
On-β	1,914	22.9	4.81

Table 5. Filter and Resampling Parameters in the Multirate Model

Synapse	M	L	R_{circ}	K	No. of Nonzero Weights	C_m	D_m	Convergence Range
Cone $\Rightarrow$ CBb_1	2	1	1	0.2	5	5.00	1.25	5
CBb_1 $\Rightarrow$ on-β	16	9	13.35	0.004	561	6.93	2.19	4–9

by the variation in weight values of the 81 space-varying filters, the largest value corresponds to the peak of the direct-form filter. Assuming a circularly symmetric, Gaussian form for the swf, we set

$$\mathrm{swf}_{CBb_1 \Rightarrow on\text{-}\beta}(r) = 33 \exp(-r^2/13^2), \tag{21}$$

were r is in micrometers. Equation (21) implies that, on average, CBb_1 bipolars give 33 or fewer contacts when they synapse exactly at the center of an on-β cell and give approximately 3.1 contacts at a 20-μm dendritic radius. To derive the direct-form filter, we substitute $L = 9$ and $T_s = 12.8$ μm into Eq. (13) and set the circular boundary as in Eq. (17), which give

$$h_{CBb_1 \Rightarrow on\text{-}\beta}(u_1, u_2) = \begin{cases} K \exp[-(u_1^2 + u_2^2)/83.6] & (u_1^2 + u_2^2)^{1/2} \le 13.35 \\ 0 & \text{otherwise} \end{cases}. \tag{22}$$

This filter is plotted on the left-hand side of Fig. 9(b).

As we will show, $h_{CBb_1 \Rightarrow on\text{-}\beta}$ is not sufficiently low pass for the prevention of aliasing. One consequence is that the space-varying filters have slightly different zero-frequency gains, and an input consisting of a constant light level will yield an output whose samples cycle among several different but close values. One can easily solve this problem by rescaling the space-varying filters so that the sum of the coefficients in each filter is the same. The rescaled filter is plotted on the left-hand side of Fig. 9(c). A comparison of Figs. 9(b) and 9(c) shows that, for both the space and frequency domains (discussed in Subsection 2.B.2), this rescaling has little effect on the direct-form filters.

Since the two synaptic steps being modeled are in series, the overall gains of the filters can be set arbitrarily. For computational convenience we set the gains so that a constant input gives a constant output of the same intensity. Thus $K = 0.2$ in Eqs. (20) and $K = 0.004$ in Eq. (22). Table 5 summarizes the parameters of this model.

As Fig. 10(a) shows, the $CBb_1 \Rightarrow$ on-β space-varying filters are of varying shape. In Fig. 10(a) four of the 81 space-varying filters are drawn, where filled diamonds indicate the on-β samples and small open circles indicate the CBb_1 samples touched by the on-β samples. To compare these filters with the on-β cell dendritic fields shown in Fig. 10(b), we rescaled the weight values, to show the number of synaptic contacts used by the model. As we can see, much of the type of variation in dendritic field shape and in the number of synaptic contacts is embodied in the multirate model. Of course, retinal anatomic variation is much greater than the variation in the multirate model, but much of the flavor of the anatomy is represented. Note that, though the contacts for each CBb_1 cell in Fig. 10(b) are spread over a small area, the model acts as if all the synapses from a CBb_1 cell occur at one point. The model does not represent the locations of the individual synapses between a CBb_1 cell and an on-β cell; rather it represents the total number of synapses between these cells.

2. Analyzing the Model and the Effects of Aliasing

We examine the model primarily in the frequency domain. Because the sampling rate in the model decreases from one array to the next, the Nyquist rate decreases in successive arrays (Table 4). To prevent aliasing, the filters must remove the high frequencies from the signal in one array before they can alias in the next array. In particular, multirate filtering theory shows that direct-form filters in systems like that in Fig. 3(c) must attenuate components above their Nyquist rate/$\max(M, L)$ to prevent aliasing.[11,12] The magnitudes of the discrete-time Fourier transforms[32] (DTFT's) of the direct-form filters are shown in Fig. 9(a) and on the right-hand sides of Figs. 9(b) and 9(c). Figure 9(a) shows DTFT cross sections for the plus-shaped and cross-shaped filters along a 0° line (ω_x axis) and along a 45° line ($\omega_x = \omega_y$). Because these filters are simply rotated and scaled versions of each other, their DTFT's are also related by rotation and scaling. The DTFT's in Figs. 9(b) and 9(c) are essentially rotationally symmetric, and we show only their 0° cross sections. While all the filters in Fig. 9 are low pass, they pass with significant magnitude frequency components above their Nyquist rate/$\max(M, L)$ cutoffs (shown by the vertical lines in the figure). High-frequency components in the input will alias in both sets of synapses and introduce spurious frequencies. Figure 11 demonstrates examples of this aliasing for the plus-shaped cone $\Rightarrow$ CBb_1 filter. In Fig. 11(a) the input i_{cone} is a sinusoid of 12.75 cycles/degree. This sinusoid aliases in the CBb_1 and on-β cell arrays to 4.37 cycles/degree. $h_{cone \Rightarrow CBb_1}$ and $h_{CBb_1 \Rightarrow on\text{-}\beta}$ attenuate this frequency component, but they are insufficient to prevent the input from being confused with 4.37-cycle/degree input. In Fig. 11(b) i_{cone} is a square wave of 6 cycles/degree. The CBb_1 image (not plotted for clarity) shows the square wave, but the on-β image is quite distorted and demonstrates aliasing of several frequency components in the square wave. In the retina irregular sampling would cause this aliasing to manifest itself partly as low-frequency sinusoids and partly as broadband

noise,[33,34] both of which lead to significant degradation of the signal.[35]

We characterize the degree of aliasing for all the frequencies with the frequency response functions in Fig. 12 (f denotes the spatial frequency in units of cycles/degree). In Fig. 12(a) the dashed curve shows the response of CBb_1 cells to frequencies below the Nyquist rate of the CBb_1 cells (8.56 cycles/degree). Components of higher frequencies are aliased into the 0–8.56-cycles/degree baseband; the solid line shows the dashed-curve response plus the aliasing. As indicated by the large difference between the line and the curve, the aliased components constitute a large part of the response of the CBb_1 cells. This difference is also large in Fig. 12(b); in fact, the situation for the on-β cells is even worse because much of the CBb_1 signal that acts as input to the $CBb_1 \Rightarrow$ on-β synapse consists of the aliased frequencies shown in Fig. 12(a). Figure 12 shows results for the plus-shaped filter. The cross-shaped filter displays a similar degree of aliasing. As we discuss in Section 3, the main anatomic reason that these filters are not sufficiently low pass is that their synapses have small convergences.

The aliasing in the CBb_1 and on-β images would be a problem in the cat retina if not for several factors, which include (1) the optic pointspread function, (2) the cone aperture, (3) cone–cone gap junctions, (4) temporal blurring from eye tremor, and (5) the low-pass nature of natural scenes. Several authors have shown that the optics and the cone aperture remove frequency components that would otherwise alias in the foveal cones of primates and humans.[34,36–39] The cone–cone gap junctions and eye tremor also low-pass filter the image. The amplitudes of spectra of natural scenes generally drop in inverse proportion to spatial frequency.[40,41] In humans the high frequencies have so little energy as to make negligible what aliasing does occur.[42] In cats it is unclear whether aliasing occurs. The analysis in Fig. 12 considers the worst-case situation of viewing a sharp, narrow line, which, in the limit of being infinitely thin, has a Fourier transform with constant magnitude along the frequency axis perpendicular to the line. The influence of these factors in the cat eye can be seen in the optics-to-cone frequency response $t_{cone}(f)$ [Fig. 13(a)], which we derive from the optics-to-cone spatial impulse response as calculated by Smith and Sterling.[5] The term optics-to-cone indicates that the response includes all spatial processing that occurs from the cornea to the cone pedicle outputs, that is, the eye's optics, the cones and the cone aperture, the

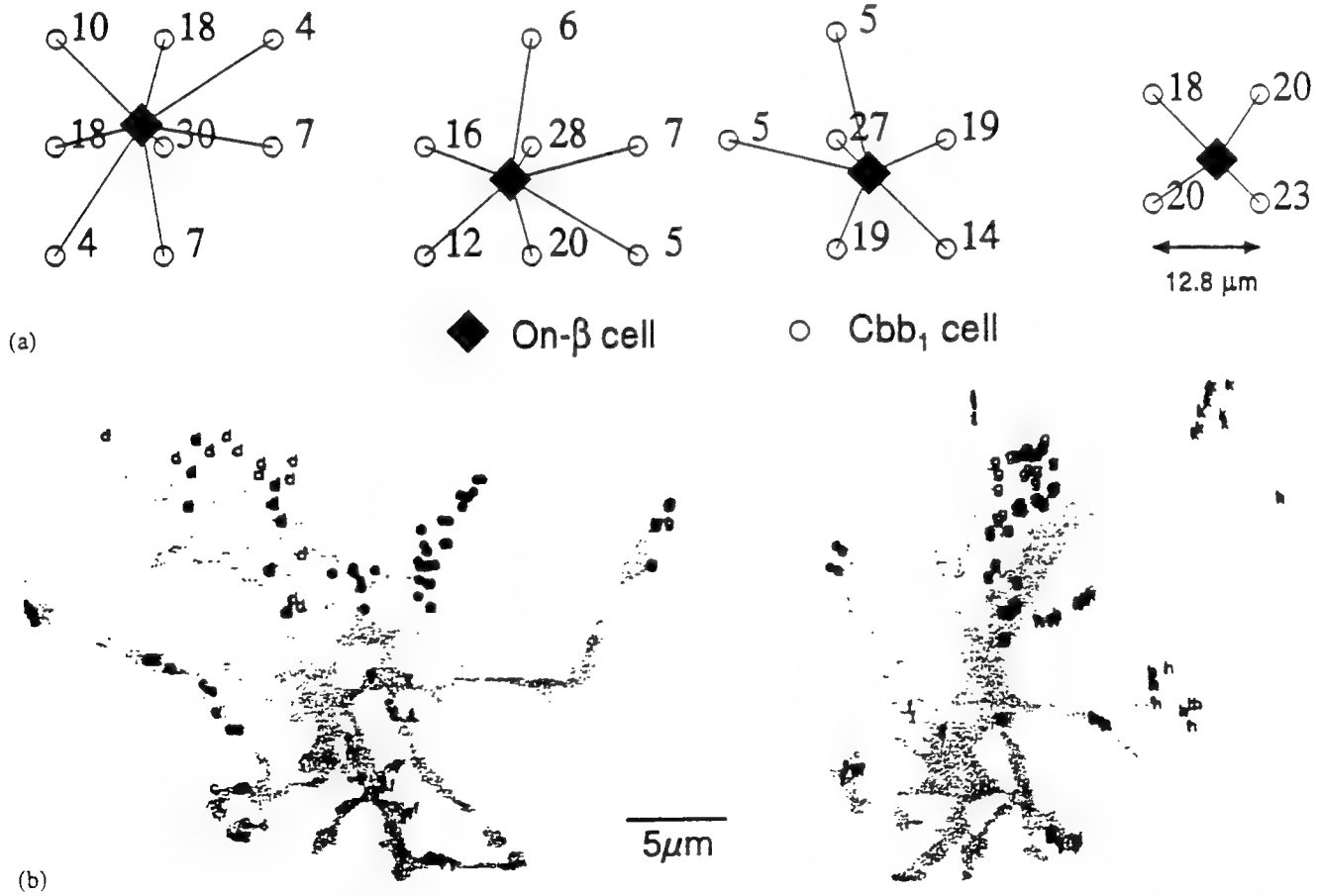


Fig. 10. Comparison of typical space-varying filters for the $CBb_1 \Rightarrow$ on-β synapse with on-β cell dendritic fields. (a) Space-varying filters g_{11}, g_{22}, g_{33}, and g_{33}. Filled diamonds indicate the on-β samples, and small open circles indicate the CBb_1 samples touched by the on-β samples. The contacts are shown explicitly by the lines. The offset of the diamonds from the circles reflects the different alignments between samples in arrays of different densities. The numbers are the filter weight values rescaled so that they show the number of synaptic contacts from each CBb_1 cell in the model. (b) Tangential projections of on-β cell dendritic trees at 1° eccentricity. Different letters refer to different CBb_1 bipolars. Each occurrence of a letter indicates a synaptic contact between that bipolar and the on-β cell (on-β cells from Ref. 6).

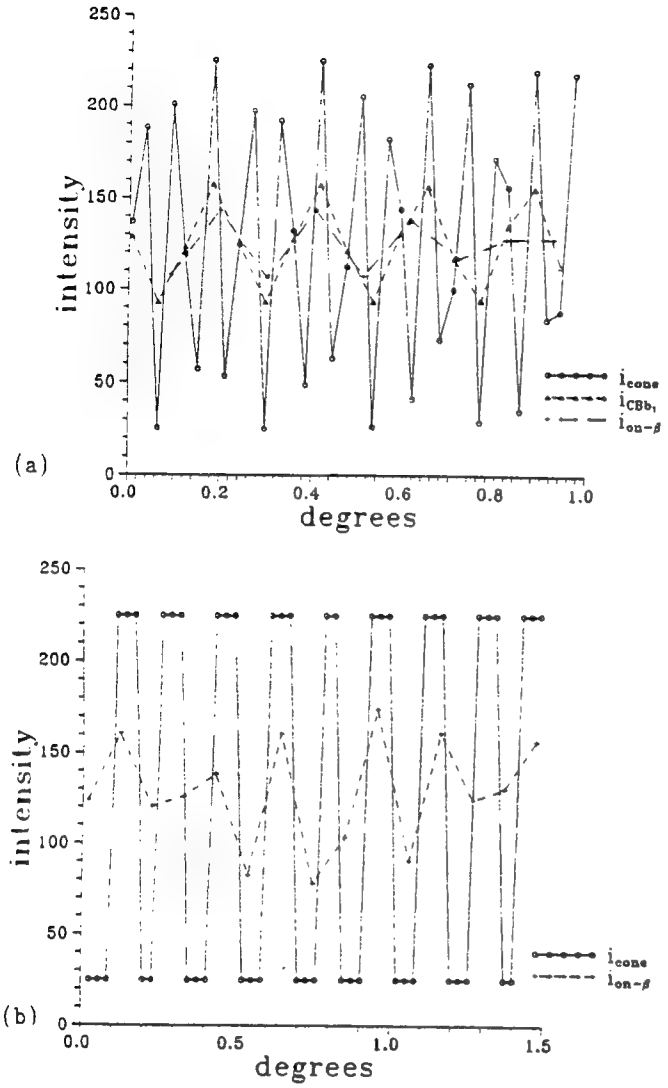


Fig. 11. Demonstrations of aliasing in spatial domain responses of CBb_1 and on-β cells in the model. The $h_{\bar{cone} \Rightarrow CBb_1}$ filter is used. The lines are not continuous signal reconstructions but simply connect adjacent samples. (a) The cone signal (○) is a sinusoid of 12.75 cycles/degree, a frequency normally severely attenuated by the optics and the cone–cone gap junctions. The aliased CBb_1 (△) and on-β (+) outputs are sinusoids at 4.37 cycles/degree and appear the same as they would if the cone image were a 4.37-cycle/degree sinusoid of lower intensity. (A sinusoid at 12.75 cycles/degree is 4.19 cycles/degree above the CBb_1 Nyquist rate of 8.56; thus it aliases to 8.56 − 4.19 = 4.37 cycles/degree.) Because of their lower sampling rates, the CBb_1 signal has one sample for every two cone samples and the on-β signal has nine samples for every 32 cone samples. (b) The cone signal (○) is a square wave of 6 cycles/degree. The on-β output (−) shows aliasing of several frequency components in the square wave. The CBb_1 signal follows the square wave; it is omitted for clarity.

cone–cone gap junctions, and the A and B horizontal cells. Smith and Sterling calculated the cat optics-to-cone spatial impulse reponse by deconvolving the on-β cell receptive field with a function based on the synaptic weighting between cones and on-β cells. Frequency components above 5.48 cycles/degree are attenuated to 1/100 of the maximum of $i_{cone}(f)$. Thus the 12.75- and 6-cycle/degree signals in Fig. 11 are lost to neural noise before they can alias.

Using the mathematics of multirate filtering, we can compute the CBb_1 and on-β cell frequency responses in terms of the optics-to-cone frequency response. The frequency response for the output of the one-dimensional system shown in Fig. 3(c) is given by

$$Y[\exp(j\omega)] = \frac{1}{M}\sum_{i=0}^{M-1} H\left[\exp\left(\frac{j\omega}{M} - \frac{j2\pi i}{M}\right)\right] \times X\left[\exp\left(\frac{j\omega}{M} - \frac{j2\pi i}{M}\right)\right], \quad (23)$$

where $H[\exp(j\omega)]$ is the DTFT of $h(u)$ and $j = \sqrt{-1}$.[11] Applying Eq. (23) to Fig. 7, substituting for L and M, and

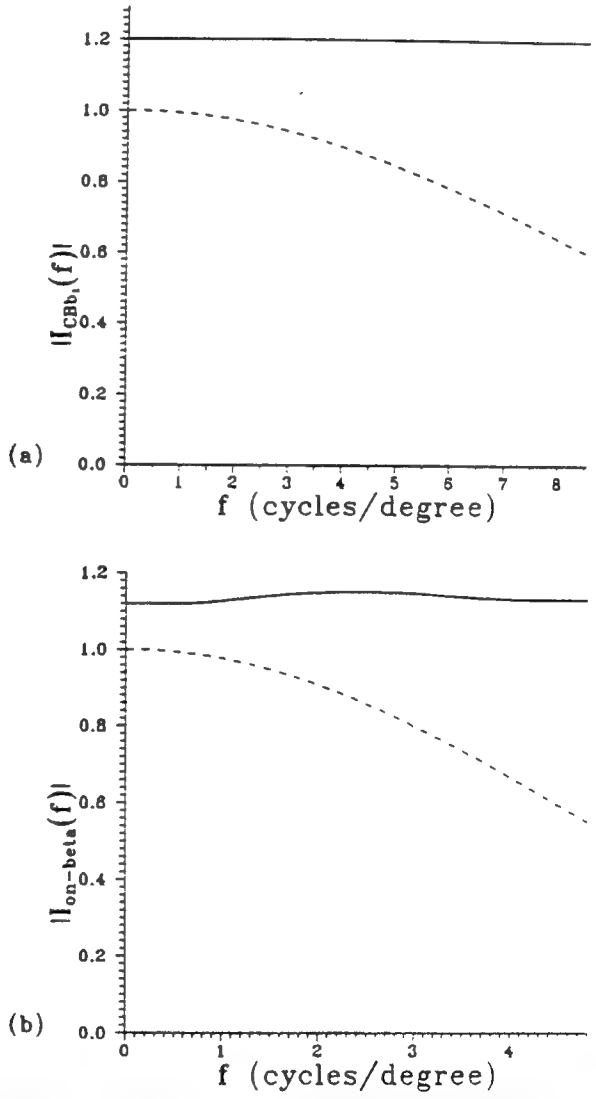


Fig. 12. Demonstration of aliasing in spatial frequency responses of CBb_1 and on-β cells in the model when there is no prefiltering. Each response is normalized so that the maximum nonaliased response equals unity. The frequency axis f is in units of cycles/degree. For each plot, f extends to the Nyquist rate for that cell array. The dashed curves show the response to input frequency components up to the Nyquist rate for that cell array. The solid lines show the dashed response plus the frequencies that alias. (a) Frequency response and aliasing of CBb_1 cell layer. (b) Frequency response and aliasing of on-β cell layer. The plotted results in (a) and (b) are for the plus-shaped filter. The cross-shaped filter demonstrates a similar degree of aliasing.

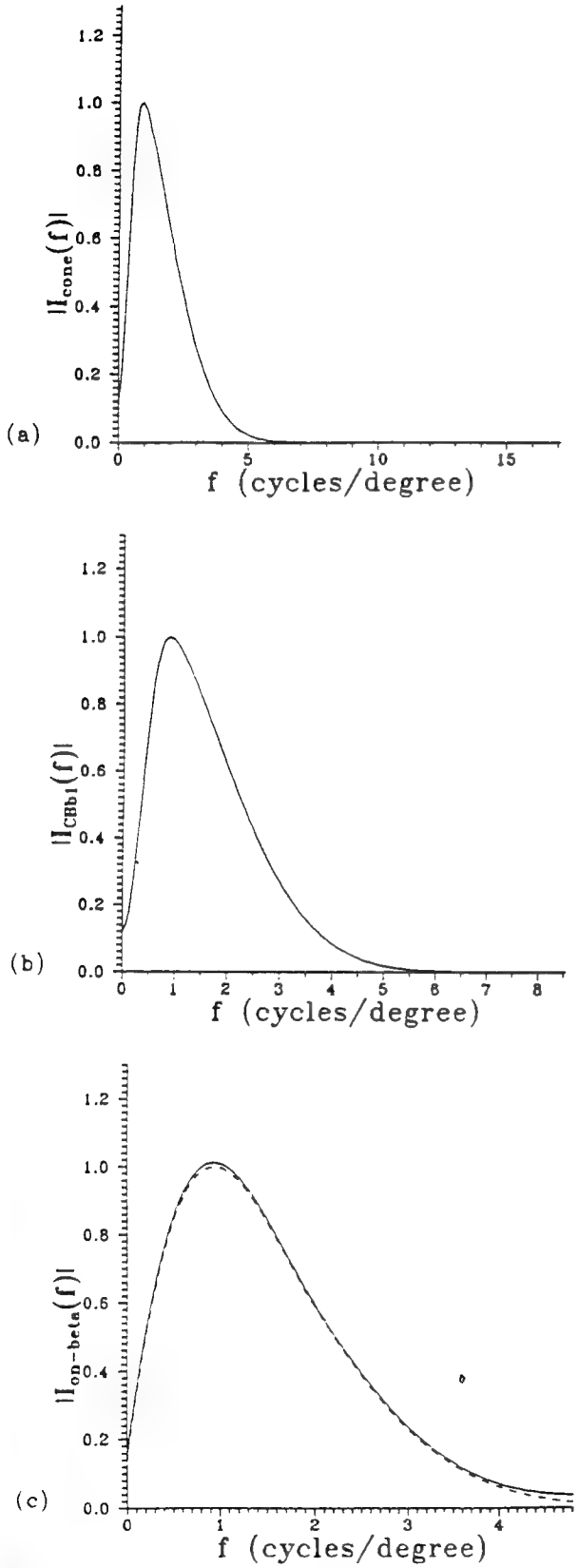


Fig. 13. Normalized spatial frequency responses in the model when prefiltering is included. (a) Cone spatial frequency response at 1° eccentricity based on computations by Smith and Sterling (balance factor $b = 0.9$).[8] (b) Frequency response of CBb_1 cell layer based on (a). Solid and dashed curves overlap. (c) Frequency response and aliasing of on-β cell layer based on (a). The results for the plus-shaped and cross-shaped filters are essentially identical for all three plots (see the text).

converting radian frequency to spatial frequency gives

$$I_{CBb_1}(f) = \frac{1}{2}\left|\sum_{i=0}^{1} H_{cone\Rightarrow CBb_1}[\exp(j40.39f - j\pi i)] \times I_{cone}[\exp(j40.39f - j\pi i)]\right|, \quad (24)$$

$$I_{on\text{-}\beta}(f) = \frac{1}{16}\left|\sum_{i=0}^{15} H_{CBb_1\Rightarrow on\text{-}\beta}\left[\exp\left(j8.976f - \frac{j\pi i}{8}\right)\right] \times I_{CBb_1}\left[\exp\left(j8.976f - \frac{j\pi i}{8}\right)\right]\right|, \quad (25)$$

where $H_{cone\Rightarrow CBb_1}[\exp(j\omega)]$ is the DTFT of $h_{cone\Rightarrow CBb_1}$, $H_{CBb_1\Rightarrow on\text{-}\beta}[\exp(j\omega)]$ is the DTFT of $h_{CBb_1\Rightarrow on\text{-}\beta}$, and $I_{CBb_1}(f)$ and $I_{on\text{-}\beta}(f)$ are the frequency responses of the CBb_1 and on-β images, respectively. Here $h_{cone\Rightarrow CBb_1}$ refers to either the plus-shaped or the cross-shaped filter. For clarity we have presented the one-dimensional equations and defined each output in terms of the previous signal in the system. The two-dimensional results are similar. These frequency responses are plotted in Figs. 13(b) and 13(c) for the plus-shaped filter.

As we can see in Fig. 13(b), the dashed and solid curves are identical for the CBb_1 image. There is no aliasing, since frequencies above the CBb_1 Nyquist rate are highly attenuated in the cone image. As Fig. 13(c) shows, there still remains a small amount of aliasing in the on-β image. Nevertheless, we regard this aliasing as insignificant for several reasons:

(1) As indicated by the small difference between the curves, the aliasing forms a nearly insignificant part of the on-β output.

(2) The gain for the aliased frequencies never exceeds 1.6% of the maximum on-β frequency response and is likely at or barely above the level of neural noise.

(3) Because the sampling of all three classes of cell is highly irregular,[1,4,43] the small amount of aliasing that may occur scatters into broadband noise.[33,34]

Thus, for broadband images, the attenuation of high frequencies performed before or at the cone level prevents aliasing that would otherwise occur in CBb_1 and on-β cells. These results are essentially identical for both the plus-shaped and cross-shaped filters. The root-mean-square difference for the images produced by the plus-shaped and cross-shaped filters is 0.0079 for the CBb_1 images and 0.0093 for the on-β images. These two filters produce such similar outputs because they are similar in those frequencies passed by the cones and differ significantly only in the frequencies attenuated by the cones [compare the attentuated region in Fig. 13(a) with the DTFT's in Fig. 9(a)].

We compare our calculations with results from cat psychophysics. Hall and Mitchell show that cats can both detect gratings and discriminate between vertical and horizontal gratings with equal ability up to between 8.5 and 9 cycles/degree.[44] Their results suggest that aliasing does not play a part in the detection of the gratings. On-β and off-β cells have the same densities, and the Nyquist rate of each class of cell at the area centralis is approximately 6.5 cycles/degree.[45] Since the Nyquist rate of both classes considered as one array is $6.5\sqrt{2} \approx 9.2$ cycles/degree,[46] Hall and Mitchell suggest that the on-β and off-β

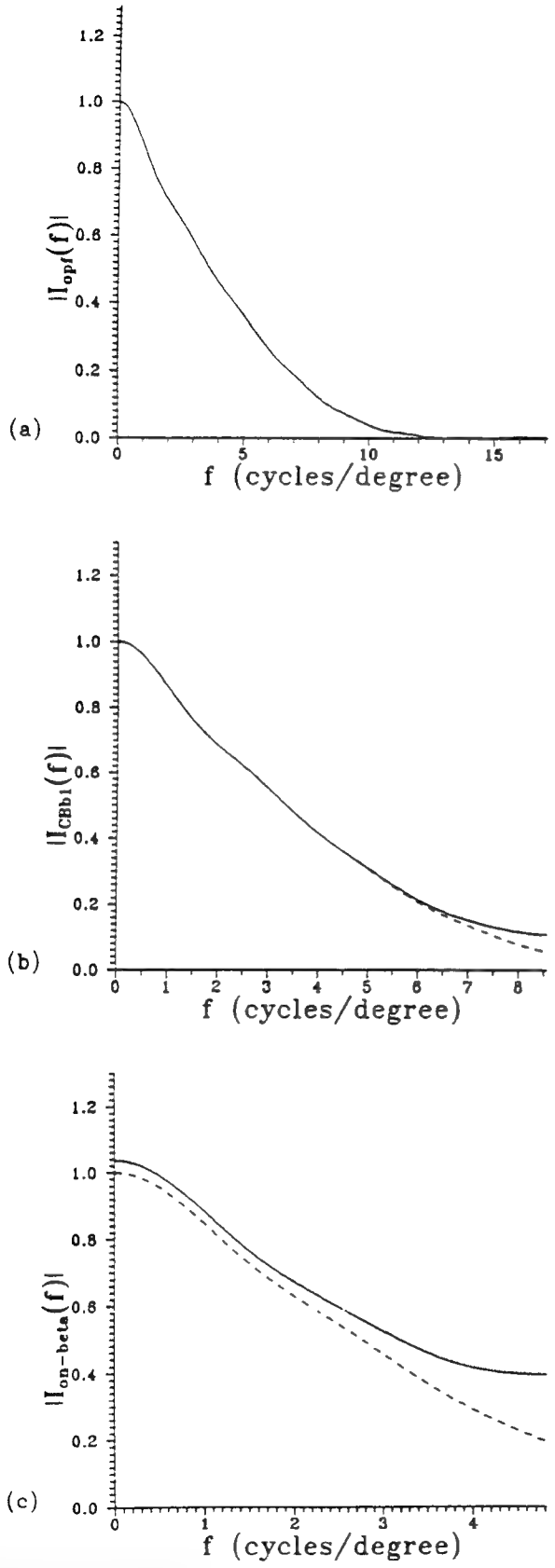


Fig. 14. Normalized spatial frequency responses in the model when prefiltering consists of only the cat optics. (a) Frequency response of cat's optics for a 3-mm pupil.[47,48] (b) Frequency response and aliasing of CBb_1 cell layer based on (a). (c) Frequency response and aliasing of on-β cell layer based on (a). The plots shown are for the plus-shaped filter. Results for the cross-shaped filter are essentially identical.

cells are interpreted by the cat brain as a single sampling mosaic.

We include both on-β and off-β cells by doubling the CBb_1 and on-β cell densities. Mathematically, this step reduces the CBb_1 and on-β cell array sampling periods by a factor of $\sqrt{2}$, scales their frequency axes in Fig. 13 by the same amount, and doubles the cone $\Rightarrow$ CBb_1 divergence to 2.4 [Eq. (12)]. The maximum grating acuity in the cat is determined by the center of the retina, the region with the highest cell densities. To compare our analysis with that in Ref. 44, we assume that the convergences and the divergences of the on-β and off-β cell pathways are the same at 0° eccentricity as they are at 1° eccentricity (where the anatomic data are available). Cone density at 0° is typically 30,000/mm^2, while in our model d_{cone} = 24,200 at 1°.[43] The assumption permits us to reinterpret the frequency axes in Fig. 13 for 0° eccentricity by rescaling the axes by 30,000/24,200. Thus the abscissa of Fig. 13(c) is rescaled from its present value of 4.81 to $(4.81)(30{,}000/24{,}200)(\sqrt{2}) = 8.43$ cycles/degree. Figure 13(c) then implies that frequency components up to approximately 8.43 cycles/degree are passed by the model with significant gain and without aliasing. Beyond this frequency, components are severely attenuated, and what aliasing remains is scattered into broadband noise. This cutoff is in agreement with Hall and Mitchell's measurement of 8.5–9 cycles/degree.

It is of interest to determine to what degree the optics of the cat eye alone are responsible for protection of the cells from aliasing. The frequency response of the cat's optics is given by Fig. 14(a).[47,48] If we substitute this response for $I_{cone}(f)$, we effectively ignore the effects of cone aperture, cone–cone gap junctions, A and B horizontal cells, and blurring by eye tremor. Figures 14(b) and 14(c) show that the optics are adequate for the removal of practically all the aliasing in the CBb_1 cells but permit significant aliasing in the on-β cells. The plotted results are for the plus-shaped filter, but they are similar to those for the cross-shaped filter. Clearly the majority of the aliasing is removed by the optics, but additional attenuation of frequencies above 4.81 cycles/degree is necessary for its complete prevention. This attenuation is largely performed by the cone aperture and the cone–cone gap junctions. Owing to the irregular sampling, much of the aliasing in the on-β cell array would scatter into broadband noise, but the significant energy of this noise would still greatly degrade the on-β signal.

3. DISCUSSION

Many anatomic properties are a consequence of the different densities of cell classes. Models based on multirate filtering can incorporate these properties quite naturally. The efficient implementations that are possible with multirate models permit rapid computation of the outputs for arbitrary inputs. The closed-form solutions in space and frequency domains provide a means of analyzing in detail how anatomic properties dictate the responses of cells. These features permit the derivation of generalized input/output relationships for arbitrary anatomic densities, convergences, divergences, and other cell properties. In this paper we developed a means for the modeling of synapses between several cell layers that uses multirate filtering,

thereby introducing the opportunity for a model of the effects of postreceptoral filtering. We discussed methods that ensure that a set of filters modeling synapses touches every sample in the presynaptic array. In applying the modeling technique to the cone ⇒ CBb_1 ⇒ on-β cell pathway, we calculated the frequency responses of the CBb_1 and on-β cells based on the cone frequency response, examined how aliasing in these cells is prevented, and compared our results with cat psychophysics. We showed that the optics of the cat eye are insufficient to prevent aliasing in these cells independently. Multirate analysis demonstrates that the highest spatial frequency that can be passed by the retina without aliasing is determined not always only by the densities of cones, bipolar cells, and ganglion cels but also by the synaptic and the dendritic weighting between these cells.

Because of the detail with which spatial anatomic information is incorporated into multirate models, analysis of these models can potentially permit the study of several retinal properties. The propagation of noise along converging synaptic paths in multiple-cell layers can be modeled by the association of noise processes with each synapse. The influence of convergence, divergence, synaptic and dendritic weighting, and synaptic gain on receptive fields can be examined by calculation of the appropriate transfer functions. Trade-offs between the number of synaptic layers and the volume of dendritic trees can be calculated as an architecture varies among parallel, hierarchic, and various hybrids. The approach to these trade-offs is reported in an early form in Refs. 13–15.

A distinct difference between the cat retina and our model is that retinal cells are usually laid out in a disordered manner, whereas the samples in the model are in a regular array. The consequences of this irregular sampling depend on whether subsequent neural processing senses the locations of the cells and uses this information.[35,49,50] Ahumada described mechanisms whereby the visual system could determine the location of its photoreceptors and alter its processing accordingly.[51] Hirsch and Miller showed that human acuity measurements are matched up to 1.5° eccentricity by primate cone nnd data scaled to the human retina without any correction for sampling disorder.[52] They conclude that cone positions are known to the visual system, at least within this range. We adopt this view for the cat retina.

When sample locations are known for a finite array of irregularly spaced samples, signal processing theory shows that perfect reconstruction of a bandlimited input is possible, regardless of the degree of irregularity, provided that there is no noise in the signal. Noise interferes with the reconstruction of signals sampled on either regular or irregular lattices; but the more the sampling deviates from a regular lattice, the more susceptible to noise is the reconstruction.[53] Thus, in Fig. 2, irregularly spaced samples of $x(n)$ can be regarded as an enhancement of the noise in $y(m)$ caused by noise in $x(n)$.

Irregular sampling also has consequences in the implementation and the design of the multirate model. We note an important distinction between irregular spacing of the photoreceptors and irregular spacing of postreceptoral cells. Photoreceptors sample a continuous light image, and their exact locations are given importance by virtue of the spatial content of the image. Postreceptoral cells can be regarded as establishing patterns of connectivity from one layer to the next. Since the cell bodies can be shifted without a change in the patterns of connections, their exact spatial positions are not overly significant. The situation is analogous in the implementation of the model. Once the filters are designed, the samples can be mathematically irregularly displaced without any effect on their values or on the filters that connect them. The situation is somewhat different for designing the model. The relative locations of samples in all the model's arrays are used in the design of the space-varying filters. While the sample locations are given importance in this manner, they can be regarded as average relative distances between the terminal branches of a presynaptic cell and the center of a postsynaptic cell's dendritic field. This approach matches the definitions of $swf(x, y)$ and $dwf(x, y)$ as average weights for class A cells whose terminal branches are located at (x, y) relative to the center of a class B cell dendritic field.

Even with regular sampling the multirate approach displays some of the variation and the irregularity seen in retinal anatomy. The irregularity is unexpected, because multirate models are based only on average measures. Yet, given different presynaptic and postsynaptic cell densities, there is no choice but that of variety in dendritic fields, even if the cells are in a regular array. Since this variation [Fig. 10(a)] occurs with period L along each axis and is not so great as anatomic variation [Fig. 10(b)], it suggests that anatomic cell-to-cell irregularity in the retina has two components: (1) variation that is due to the different relative locations of cells in presynaptic and postsynaptic cell arrays as required by different cell densities and (2) variation that is due to partly random branching, direction, and length of cell growth.

Multirate filtering provides a means of examining the amplification or the aliasing of frequency components that propagate through cell layers of different densities. For example, for the case of resampling as in Fig. 3(c), if the filter has the appropriate cutoff, then the maximum frequency that can be passed by the system is the lower of the two Nyquist rates of the input $x(n)$ and the output $y(m)$. If the filter has a higher cutoff, as is the case in the cone ⇒ CBb_1 ⇒ on-β example, components in $x(n)$ will alias in $y(m)$ unless they are attenuated before $x(n)$. If the filter has a lower cutoff, a smaller range of frequencies is passed to $y(m)$. Thus the highest frequency passed without aliasing is determined not only by the densities of the cell layers but also by the swf and the dwf between the layers. The present analysis suggests that the attenuation of high frequencies in the cones prevents aliasing that would otherwise occur in CBb_1 and on-β cells. The analysis also suggests that, while assuming responsibility for most of this attenuation, the cat's optics cannot independently prevent all the aliasing in these cells.

The two filters in the cone ⇒ CBb_1 ⇒ on-β model permit aliasing primarily because their convergences are so small. As we showed in Subsection 2.A.2, convergence corresponds to the average number of weights in the space-varying filters. The fewer weights there are in a low-pass filter, the less well the filter approximates an ideal low-pass filter and the more it passes frequency components above the ideal cutoff. Two-dimensional space-varying filters with only five to seven weights cannot possibly act

sufficiently low pass for the prevention of most of the aliasing in the model, and the prefiltering in the cone image is necessary. In this regard, it is of interest that the convergences of the cone ⇒ A and the cone ⇒ B horizontal cell synapses near the area centralis in the cat retina are of the order of 140 and 90, respectively.[1] The ratios of presynaptic to postsynaptic cell densities between cones and either type of horizontal cell are much greater than those for either of the synapses in the cone ⇒ CBb_1 ⇒ on-β pathway.[2] These greater density ratios provide the potential for much more aliasing than that in our example, even with the cone prefiltering. The large convergences may be necessary for preventing the aliasing associated with low horizontal cell densities.

ACKNOWLEDGMENTS

We thank Peter Sterling for many helpful discussions and for his help with the anatomical pictures. This study was supported by U.S. Air Force Office of Scientific Research grant 91-0082. B. Levitan was also supported by a National Institutes of Health fellowship under grant 5-T32-GM07170.

REFERENCES

1. H. Wässle, B. B. Boycott, and L. Peichl, "Receptor contacts of horizontal cells in the retina of the domestic cat," Proc. R. Soc. London Ser. B **203**, 247–267 (1978).
2. H. Wässle, L. Peichl, and B. B. Boycott, "Topography of horizontal cells in the retina of the domestic cat," Proc. R. Soc. London Ser. B **203**, 269–291 (1978).
3. E. Cohen and P. Sterling, "Microcircuitry related to the receptive field center of the on-beta ganglion cell," J. Neurophysiol. **65**, 352–359 (1991).
4. E. Cohen and P. Sterling, "Demonstration of cell types among cone bipolar neurons of cat retina," Philos. Trans. R. Soc. London Ser. B **330**, 305–321 (1990).
5. E. Cohen and P. Sterling, "Convergence and divergence of cones onto bipolar cells in the central area of cat retina," Philos. Trans. R. Soc. London Ser. B **330**, 323–328 (1990).
6. E. Cohen and P. Sterling, "Parallel circuits from cones to the on-beta ganglion cell," Eur. J. Neurosci. **4**, 506–520 (1992).
7. M. A. Freed, R. G. Smith, and P. Sterling, "Computational model of the on-alpha ganglion cell receptive field based on bipolar cell circuitry," Proc. Natl. Acad. Sci. USA **89**, 236–240 (1992).
8. R. G. Smith and P. Sterling, "Cone receptive field in cat retina computed from microcircuitry," Visual Neurosci. **5**, 453–461 (1990).
9. Y. Tsukamoto and P. Sterling, "Spatial summation by ganglion cells: some consequences for the efficient encoding of natural scenes," Neurosci. Res. Suppl. **15**, S185–S198 (1991).
10. P. P. Vaidyanathan, "Multirate digital filters, filter banks, polyphase networks, and applications: a tutorial," Proc. IEEE **78**, 56–93 (1990).
11. R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, N.J., 1983).
12. P. P. Vaidyanathan, *Multirate Systems and Filter Banks* (Prentice-Hall, Englewood Cliffs, N.J., 1993).
13. B. Levitan and G. Buchsbaum, "A multirate filter bank perspective of early vision," in *Visual Communications and Image Processing IV*, W. A. Pearlman, ed., Proc. Soc. Photo-opt. Instrum. Eng. **1199**, 1069–1085 (1989).
14. B. Levitan and G. Buchsbaum, "Parallel, hierarchic and hybrid processing tradeoffs in early vision," Invest. Ophthal. Vis. Sci. **31**, 88 (1990).
15. B. Levitan and G. Buchsbaum, "A multirate filter bank perspective of retinal processing," in *Applied Vision*, Vol. 16 of 1989 OSA Technical Digest Series (Optical Society of America, Washington, D.C., 1989), pp. 9–12.
16. C. Enroth-Cugell and P. Lennie, "The control of retinal ganglion cell discharge by receptive field surrounds," J. Physiol. (London) **247**, 551–578 (1975).
17. C. Enroth-Cugell, J. G. Robson, D. E. Schweitzer-Tong, and A. B. Watson, "Spatio-temporal interactions in cat retinal ganglion cells showing linear spatial summation," J. Physiol. (London) **341**, 279–307 (1983).
18. C. Enroth-Cugell and J. G. Robson, "The contrast sensitivity of retinal ganglion cells of the cat," J. Physiol. (London) **187**, 517–552 (1966).
19. B. Levitan and G. Buchsbaum, "Conversions between parallel and hierarchic architecture analysis multirate filter banks," IEEE Trans. Signal Process. **40**, 2837–2841 (1992).
20. R. G. Smith and P. Sterling, "Functional consequences of morphology in types A and B horizontal cells of the cat retina," Soc. Neurosci. Abstr. **19**, 1012 (1991).
21. R. S. Shapley and C. Enroth-Cugell, "Visual adaptation and retinal gain controls," Prog. Retinal Res. **3**, 263–343 (1984).
22. M. J. M. Lankheet, R. J. A. van Wezel, and W. A. van de Grind, "Light adaptation and frequency transfer properties of cat horizontal cells," Vision Res. **31**, 1129–1142 (1991).
23. T. N. Cornsweet and J. I. Yellott, "Intensity-dependent spatial summation," J. Opt. Soc. Am. A **2**, 1769–1786 (1985).
24. M. V. Srinivasan, S. B. Laughlin, and A. Dubs, "Predictive coding: a fresh view of inhibition in the retina," Proc. R. Soc. London Ser. B **216**, 427–459 (1982).
25. A. M. Rohaly and G. Buchsbaum, "Global spatiochromatic mechanism accounting for luminance variations in contrast sensitivity functions," J. Opt. Soc. Am. A **6**, 312–317 (1989).
26. J. I. Yellott, "Photon noise and constant-volume operators," J. Opt. Soc. Am. A **4**, 2418–2446 (1987).
27. M. A. Freed, R. G. Smith, and P. Sterling, "The rod bipolar array in cat retina: pattern of input from rods and GABA-accumulating cells," J. Comp. Neurol. **266**, 445–455 (1987).
28. J. Kovačević and M. Vetterli, "Non-separable multidimensional perfect reconstruction filter banks and wavelet bases," IEEE Trans. Inf. Theory **38**, 533–555 (1992).
29. E. Viscito and J. P. Allebach, "The analysis and design of multidimensional FIR perfect reconstruction filter banks for arbitrary sampling lattices," IEEE Trans. Circuits Syst. **38**, 29–41 (1991).
30. E. D. Cohen, "Convergence of cone input through cone bipolar array to individual on-beta ganglion cells," thesis dissertation (Department of Anatomy, University of Pennsylvania, Philadelphia, Pa., 1987).
31. C. A. Curcio and K. R. Sloan, "Packing geometry of human cone photoreceptors: variation with eccentricity and evidence for local anisotropy," Visual Neurosci. **9**, 169–180 (1992).
32. A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, N.J., 1989), p. 45.
33. J. I. Yellott, "Spectral analysis of spatial sampling by photoreceptors: topological disorder prevents aliasing," Vision Res **22**, 1205–1210 (1982).
34. J. I. Yellott, "Image sampling properties of photoreceptors: a reply to Miller and Bernard," Vision Res. **24**, 281–282 (1984).
35. T. R. J. Bossomaier, A. W. Snyder, and D. G. Stavenga, "Irregularity and aliasing: solution?" Vision Res. **25**, 145–147 (1985).
36. D. R. Williams, "Aliasing in human foveal vision," Vision Res **25**, 195–205 (1985).
37. A. W. Snyder and W. H. Miller, "Photoreceptor diameter and spacing for highest resolving power," J. Opt. Soc. Am. **67**, 696–698 (1977).
38. W. H. Miller and G. D. Bernard, "Averaging over the foveal receptor aperture curtails aliasing," Vision Res. **23**, 1365–1369 (1983).
39. D. I. A. Macleod, D. R. Williams, and W. Makous, "A visual nonlinearity fed by single cones," Vision Res. **32**, 347–364 (1992).
40. G. J. Burton and I. R. Moorehead, "Color and spatial structure in natural scenes," Appl. Opt. **26**, 157–170 (1987).
41. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," J. Opt. Soc. Am. A **4**, 2379–2394 (1987).
42. S. J. Galvin and D. R. Williams, "No aliasing at edges in normal viewing," Vision Res. **32**, 2251–2259 (1992).

43. R. H. Steinberg, M. Reid, and P. L. Lacy, "The distribution of rods and cones in the retina of the cat," J. Comp. Neurol. **148**, 229–248 (1973).
44. S. E. Hall and D. E. Mitchell, "Grating acuity of cats measured with detection and discrimination tasks," Behav. Brain Res. **44**, 1–9 (1991).
45. H. Wässle, L. Peichl, and B. B. Boycott, "A spatial analysis of on- and off-ganglion cells in the cat retina," Vision Res. **10**, 1151–1160 (1983).
46. A. Hughes, "Cat retina and the sampling theorem: the relation of transient and sustained brisk-unit cut-off frequency to the α- and β-mode density," Exp. Brain Res. **42**, 196–202 (1981).
47. J. G. Robson and C. Enroth-Cugell, "Light distribution in the cat's retinal image," Vision Res. **18**, 159–173 (1978).
48. H. Wässle, "Optical quality of the cat eye," Vision Res. **11**, 995–1006 (1971).
49. A. Papoulis, "Error analysis in sampling theory," Proc. IEEE **54**, 947–955 (1966).
50. A. S. French, A. W. Snyder, and D. G. Stavenga, "Image degradation by an irregular retinal mosaic," Biol. Cybernet. **27**, 229–233 (1977).
51. A. J. Ahumada, Jr., "Learning receptor positions," in *Computational Models of Visual Processing*, M. S. Landy and J. A. Movshon, eds. (MIT Press, Cambridge, Mass., 1991).
52. J. Hirsch and W. H. Miller, "Does cone positional disorder limit resolution?" J. Opt. Soc. Am. A **4**, 1481–1492 (1987).
53. J. L. Yen, "On nonuniform sampling of bandwidth-limited signals," IRE Trans. Circuit Theory **3**, 251–257 (1956).

Conversions Between Parallel and Hierarchic Architecture Analysis Multirate Filter Banks

Bennett Levitan and Gershon Buchsbaum

Abstract—**We derive general conversions between equivalent parallel and hierarchic analysis multirate filter banks (MRB's) as well as sufficient conditions for existence and uniqueness of the conversions. We use MRB's with arbitrary, rational number changes in sampling rate between successive outputs and arbitrary LTI filtering for each output. Conversion consists of commuting sampling rate expanders, sampling rate compressors, and filters to turn one MRB into the form of the other. For a class of MRB's we call "well-formed," the conversions between architectures are one to one.**

I. Introduction

An analysis multirate filter bank (MRB) consists of a set of filters that produces several different sampling rate, or spatial scale, versions of a signal. MRB's can operate very efficiently by allowing differential allocation of processing and storage resources to the spatial scales [2], [6], [11], [14]–[17], [19], [20]; however, these advantages depend heavily on the architecture. In a hierarchic architecture, the nth output is computed by operations on the n-1st output; the first output is computed directly from the input. In a parallel architecture, all outputs are computed directly from the input. If the corresponding outputs of two MRB's (with the same or different architectures) are equal for all inputs, the MRB's are *equivalent*. Compared to a parallel architecture, an equivalent hierarchy has the advantages of: i) smaller filter sizes, ii) lower sampling rates for most computation, and iii) a smaller total number of connections between processing elements, a feature important for "hardwired" implementations [1], [5], [9], [12], [15]. The hierarchy's disadvantages include: i) slower operation in hardwired implementations, ii) less flexibility in choosing filters, iii) less straightforward design, and iv) susceptibility to noise and errors propagating through successive outputs.

To benefit from these and other tradeoffs between equivalent hierarchic and parallel analysis MRB's requires a method to convert between the parallel and hierarchic architectures and means to determine when such conversions are allowed. These issues have been examined for some MRB's [2]–[4], [12], [18] but not for the general case. In this correspondence, we develop equations and conditions for the conversions.

II. Theory

A. Definitions of Multirate Filter Bank Architectures

In the parallel multirate filter bank, each output y_n (x_n) is computed from the original input $y_0(x_0)$ in three steps: i) sampling rate expansion by factor $\boldsymbol{L}_n$, ii) filtering by arbitrary, linear time-invariant (LTI) parallel filter $\boldsymbol{h}_n(\boldsymbol{u}_n)$, and iii) sampling rate compression by factor $\boldsymbol{M}_n$. These basic operations of multirate filtering are described extensively in [5], [17]. Fig. 1(a) shows the parallel MRB

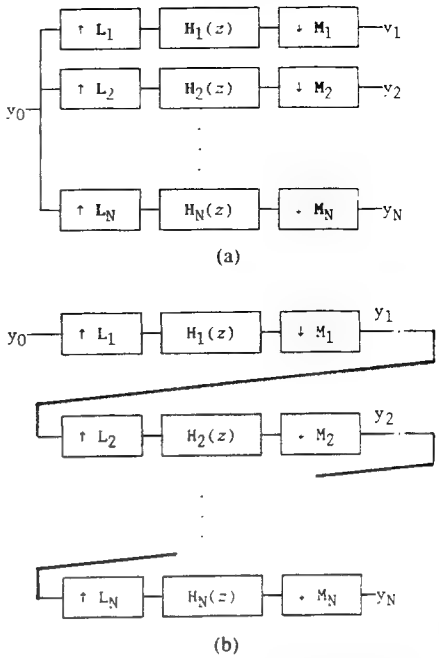


Fig. 1. Multirate filter bank architectures. (a) Parallel MRB: Output y_n is produced by expanding, filtering, and compressing input y_0. $\uparrow L_i$ and $\downarrow M_i$ indicate sampling rate expansion and compression by L_i and M_i, respectively. (b) Hierarchic MRB: Output y_n is produced by expanding, filtering, and compressing y_{n-1}. Parallel MRB variables are boldface versions of the corresponding hierarchic MRB variables. The spatial variables for the filters and outputs are indexed to differentiate between coordinates on different levels.

in the z domain, where $\boldsymbol{H}_n(z)$ is the z transform of $\boldsymbol{h}_n(\boldsymbol{u}_n)$. To distinguish between parallel and hierarchic MRB variables, all parallel MRB variables are in boldface type.

The hierarchic multirate filter bank successively produces outputs $y_1(x_1)$ through $y_N(x_N)$ in N stages. The nth stage computes y_n from y_{n-1} in three steps: i) sampling rate expansion by factor L_n, ii) filtering by arbitrary, LTI hierarchic filter $h_n(u_n)$, and iii) sampling rate compression by factor M_n. Fig. 1(b) shows the hierarchic MRB in the z domain, where $H_n(z)$ is the z transform of $h_n(u_n)$.

The sets of parallel and hierarchic MRB's can be partitioned into equivalence classes [10]. Each parallel equivalence class contains an infinite number of equivalent MRB's whose resampling factors satisfy $\boldsymbol{M}_n/\boldsymbol{L}_n = \boldsymbol{K}_n$, where $\boldsymbol{K}_n$ is a rational number. From elementary properties of prime numbers, it follows that each class contains one and only one MRB whose resampling factors satisfy

$$\boldsymbol{M}_n \text{ and } \boldsymbol{L}_n \text{ are relatively prime (have no common divisor other than one)}, \quad n = 1 \cdots N. \tag{1}$$

Similarly, each hierarchic MRB class contains an infinite number of equivalent MRB's whose resampling factors satisfy $M_n/L_n = K_n$, where K_n is a rational number. In each class there is one and only one MRB whose resampling factors satisfy

$$M_n \text{ and } L_n \text{ are relatively prime}, \quad n = 1 \cdots N. \tag{2}$$

We call MRB's satisfying (1) or (2) "well formed." As shown in the next section, compared to equivalent non-well-formed MRB's, well-formed MRB's are more likely to satisfy the conversion conditions derived below. Because each equivalence class is uniquely identified by one well-formed MRB, *for well-formed hier-*

Manuscript received April 5, 1990; revised September 6, 1991. This work was supported in part by AFOSR Grant 91-0082. The work of B. Levitan was also supported by an NIH fellowship under Grant 5-T32-GM07170.

The authors are with the Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA 19104-6392.

IEEE Log Number 9202795.

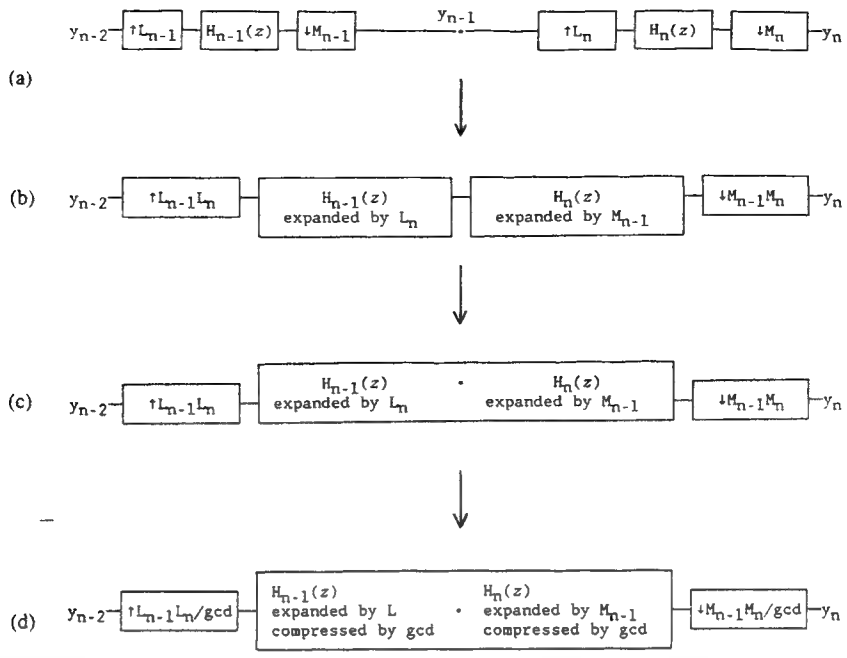


Fig. 2. Adjacent-pair commutation. The expanders and compressors of adjacent stages are commuted to form a single stage. (a) Two hierarchic stages form output y_n from y_{n-2}. (b) The result after commuting expander L_n to the left and compressor M_{n-1} to the right. We assume L_n and M_{n-1} are relatively prime. (c) The result after the filters are joined. (d) The final result after gcd, the greatest common divisor of $L_{n-1} \cdot L_n$ and $M_{n-1} \cdot M_n$ is removed from the factors.

archic and parallel MRB's that satisfy the conversion conditions, the conversions are one to one.

B. Relations between Multirate Filter Bank Architectures

1) Conversion from Hierarchic MRB to Parallel MRB:

a) General conversion: The basic step for MRB conversion is the "adjacent pair" commutation (Fig. 2). Adjacent pair commutations convert two adjacent stages into a single equivalent stage. Using rules for commuting multirate operations with filters [4] (these rules are also referred to as the "noble identities" [17]), the L_n-fold expander is commuted to the left and combined with the first expander. Similarly, the M_{n-1}-fold compressor is commuted to the right and combined with the second compressor (Fig. 2(b)). The filter system functions are expanded by the factors of the resamplers commuted with their filters. The filters are then joined (Fig. 2c). If the resulting resampling factors are not relatively prime, their greatest common divisor can be removed if the combined filter is also downsampled by the common divisor (Fig. 2d). This method is similar to that used in [4], [7].

Adjacent pair commutations are not always possible, since an L-fold expander and an M-fold compressor can commute if and only if L and M are relatively prime [7], [17]. By removing the greatest common divisor of the resampling factors in adjacent pair commutations, the reduced factors are more likely to be relatively prime to the factors of other stages during subsequent commutations. For this reason, removing common divisors in each stage by converting a hierarchic MRB to its well-formed equivalent [10] increases the likelihood of its conversion.

$n-1$ adjacent pair commutations turn the first n stages of a hierarchic MRB into a single stage computing output y_n directly from y_0. The stages can be commuted in any sequence. The set of relatively prime conditions for a particular commutation sequence is the "hierarchic to parallel commutation condition" for that sequence. If a particular sequence requires commuting factors that are not relatively prime, conversion in that sequence is not possible. *A hierarchic MRB can be converted into an equivalent parallel MRB if at least one sequence of commutations is possible for each output.* The final adjacent pair commutation in a sequence insures that M_n and L_n are relatively prime. *Thus, the parallel MRB derived from a hierarchic MRB satisfies (1) and is well formed.*

Conversion with any sequence of commutation gives equivalent parallel factors

$$L_n = \prod_{i=1}^{n} L_i / \gcd\left(\prod_{i=1}^{n} L_i, \prod_{i=1}^{n} M_i\right) \qquad n = 1 \cdots N \tag{3}$$

and

$$M_n = \prod_{i=1}^{n} M_i / \gcd\left(\prod_{i=1}^{n} L_i, \prod_{i=1}^{n} M_i\right) \qquad n = 1 \cdots N \tag{4}$$

where gcd (A, B) denotes the greatest common divisor of A and B [10]. The equivalent $H_n(z)$ is the product of multiply resampled hierarchic filters H_1 through H_n and depends on which of the potential $(n-1)!$ possible sequences of conversion is used. Each sequence has a different commutation condition and removes common divisors in a different manner.

b) Relatively prime factors case: We can derive closed-form equivalent filters for a less general case. Let all resampling factors commuted in the conversion be relatively prime; that is,

$$M_i \text{ and } L_j \text{ are relatively prime,} \qquad i = 1 \cdots N, j = i + 1 \cdots N. \tag{5}$$

For example, $M_n = 3$ and $L_n = 2$ for all n. This condition sets all gcd terms in adjacent pair commutations (Fig. 2) to one making all sequences of conversion possible. Fig. 3 shows the conversions for outputs two and three in the relatively prime factors case where $H_1(z^{L_2})$ is the z transform of $h_1(x_1)$ expanded by L_2. For arbitrary

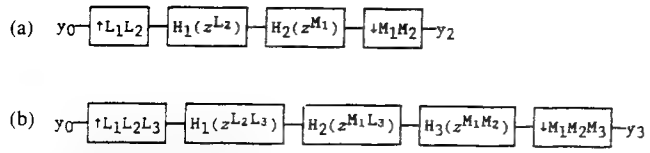


Fig. 3. Conversion of hierarchic MRB into parallel MRB for relatively prime factors case. (a) Result of using z-domain identities on Fig. 1(b) for y_2. (b) Result for y_3.

n, the parallel MRB can be written in terms of the hierarchic MRB as

$$M_n = \prod_{i=1}^{n} M_i, \quad L_n = \prod_{i=1}^{n} L_i \qquad n = 1 \cdots N \tag{6,7}$$

$$H_n(z) = \prod_{j=1}^{n} H_j(z^{W_{j,n}}) \qquad n = 1 \cdots N. \tag{8}$$

where

$$W_{1,1} = 1, \quad W_{j,n} = \prod_{i=1}^{j-1} M_i \cdot \prod_{i=j+1}^{n} L_j$$

$$n = 2 \cdots N, j = 1 \cdots n. \tag{9}$$

Equations (6) and (7) result from dropping the gcd terms in (3) and (4). Equation (8) states that the z-domain filter for level n of the equivalent parallel MRB is the product of expanded versions of hierarchic filters H_1 through H_n. It is valid for all sequences of conversion since, in any sequence, filter H_i commutes with all expanders to its right and all compressors to its left. $W_{j,n}$ is the product of factors of all expanders and compressors that commute with H_i.

The relatively prime factors case is often satisfied in practice. The sampling scheme most often used in hierarchic MRB's is reducing the sampling rate by an integer factor from y_{n-1} to y_n for all n [1], [2], [11], [12], [15], [16], [19]. This scheme corresponds to setting M_n = constant and $L_n = 1$ for all n and trivially satisfies (5).

2) Conversion from Parallel MRB to Hierarchic MRB: The parallel MRB definitions of y_{n-1} and y_n precisely define the nth hierarchic stage. To convert a parallel MRB into a hierarchic MRB, consider a system in which y_{n-1} is calculated by the parallel MRB and then filtered by a hierarchic stage to yield y_n (Fig. 4(a)). We perform an adjacent pair commutation on Fig. 4(a), but do not remove gcd $(L_{n-1} \cdot L_n, M_{n-1} \cdot M_n)$ (Fig. 4(b)). This commutation is possible only if L_n and M_{n-1} are relatively prime. To increase the likelihood of conversion, many potential common divisors can be removed by making the parallel MRB well formed before the commutations. The resulting stage (Fig. 4(b)) is similar to the parallel MRB for y_n (Fig. 4(c)); however, because it may have nonrelatively prime resampling factors, they cannot be directly compared. To compare the two, we insert C_n-fold expanders and compressors in the parallel MRB (Fig. 4(d)) and commute them to get a stage whose coefficients may not be relatively prime (Fig. 4(e)). Comparing Figs. 4(b) and (e) gives

$$L_1 = L_1, \quad L_n = C_n \cdot L_n / L_{n-1} \qquad n = 2 \cdots N \tag{10}$$

$$M_1 = M_1, \quad M_n = C_n \cdot M_n / M_{n-1} \qquad n = 2 \cdots N \tag{11}$$

$$H_1(z) = H_1(z), \quad H_n(z) = \frac{H_n(z^{C_n/M_{n-1}})}{H_{n-1}(z^{L_n/M_{n-1}})} \qquad n = 2 \cdots N. \tag{12}$$

We can write the "parallel to hierarchic commutation condition" as: *A parallel MRB can be converted into an equivalent hierarchic MRB if all M_{n-1} and L_n are relatively prime for $n = 2 \cdots N$.* L_n is calculated from the given parallel resampling factors using (10). By setting

$$C_n = L_{n-1} \cdot M_{n-1} / [\text{gcd}\,(L_n, L_{n-1}) \cdot \text{gcd}\,(M_n, M_{n-1})] \tag{13}$$

M_n and L_n can be made relatively prime. *Thus, the hierarchic MRB derived from a parallel MRB satisfies (2) and is well formed.*

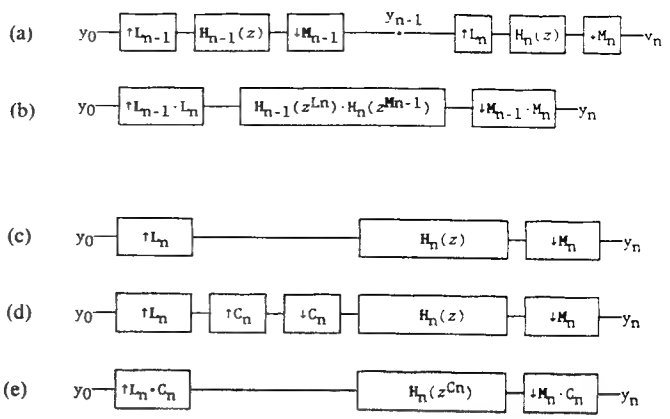


Fig. 4. Conversion of parallel MRB into a hierarchic MRB. (a) Calculation of y_n in two steps: The parallel MRB calculates y_{n-1}; a hierarchic stage calculates y_n from y_{n-1}. (b) Result after performing an adjacent pair commutation on (a) without removing the resampling factors' greatest common divisor. We assume L_n and M_{n-1} are relatively prime. (c) Well-formed parallel MRB. (d) Insertion of C_n-fold expander and compressor. Together, they do not affect the output. (e) Result after commuting and joining the resamplers.

A second condition arises from the practical requirement that the filters $H_n(z)$ in (12) be stable. $H_n(z)$ is stable provided its ROC includes the unit circle. The given parallel filters $H_n(z)$ are stable. Hence, $H_n(z^{C_n/M_{n-1}})$ and $H_{n-1}(z^{L_n/M_{n-1}})$ in (12) are also stable, because raising z to a constant does not cause poles or zeros to move on or off the unit circle. However, $H_n(z) = H_n(z^{C_n/M_{n-1}})/H_{n-1}(z^{L_n/M_{n-1}})$ may still not be stable, since $H_{n-1}0(z^{L_n/M_{n-1}})$ may have zeros on the unit circle which would exclude the unit circle from $H_n(z)$'s ROC. Hence, *A parallel MRB can be converted into a hierarchic MRB with stable filters if and only if the ratios of parallel filters $H_n(z^{C_n/M_{n-1}})/H_{n-1}(z^{L_n/M_{n-1}})$ are stable for $2 \le n \le N$.*

On a more intuitive level, this condition requires that any frequency component required in output y_n of the parallel MRB be retained in output y_{n-1}. Clearly, this component need also be retained in outputs y_1 through y_{n-2}. This point highlights an inherent difference between the two architectures: The nth filter in a hierarchy of stable filters has access only to frequency components represented in signal y_{n-1}, while parallel MRB filters have access to all components in the input y_0.

In practically every MRB we have encountered in the hierarchic processing literature, the outputs are indexed in order of decreasing sampling rate. In such MRB's, the filters $H_n(\omega)$ are nonzero for $0 \le |\omega| \le \pi/\max(L_n, M_n)$ and negligible at higher frequencies. These filters satisfy the stability condition, because the cut-offs decrease sufficiently from level $n-1$ to n. In the subband coding literature, however, filters are generally not of this form and will often not satisfy the stability condition. Subband coding analysis MRB's generally cannot be converted to a hierarchic equivalent; however, they could be implemented with difference-pyramid type hierarchies [1], [2], [12], [14], [15]. For MRB's with at least one output whose sampling rate is above that of y_0, the stability condition does not apply to any ω such that $\pi \cdot M_n/L_n < |\omega| \le \pi$.

Since these frequency components cannot be represented in y_0, they always contribute nothing to y_n.

III. Scaled Parallel Filters

Many MRB applications use scaled, or self-similar, parallel filters [1], [2], [11], [12], [19]. These filters are of the same shape but scaled differently in height and width on each level. To incorporate scaled filters into an MRB, we define the parallel filters as scaled, sampled versions of function $f_c(t)$ where t is a continuous variable:

$$h_n(u_n) = A_n \cdot f_c(u_n/B_n), \qquad n = 1 \cdots N. \tag{14}$$

We assume the MRB's satisfy the commutation conditions. A_n and B_n are positive, real-valued coefficients that allow arbitrary vertical and horizontal scaling of f_c. f_c must have a continuous argument since the B_n are arbitrary. If we define discrete function $f_n(x) = f_c(x/B_n)$ and $F_n(z)$ as its z transform, the z transform of (14) can be written

$$H_n(z) = A_n \cdot F_n(z), \qquad n = 1 \cdots N. \tag{15}$$

To solve the hierarchic filters, we substitute (15) into (12) yielding

$$H_n(z) = \frac{A_n \cdot F_n(z^{C_n/M_{n-1}})}{A_{n-1} \cdot F_{n-1}(z^{L_n/M_{n-1}})}, \qquad n = 2 \cdots N. \tag{16}$$

Equation (16) gives the hierarchic filters for scaled parallel filters based on $f_c(t)$.

VI. Discussion

This correspondence presents relations between parallel and hierarchic implementations of analysis MRB's. Fig. 5 summarizes the theory. The theory gives closed-form equations and conditions for converting a parallel MRB into an equivalent well-formed hierarchic MRB with stable filters. It gives a method and conditions to convert a hierarchic MRB into an equivalent well-formed parallel MRB and closed-form equations for the relatively prime factors case. Since MRB's of either architecture are members of equivalence classes containing an infinite number of MRB's, the conversions are obviously not unique. However, for well-formed MRB's satisfying the conversion conditions, conversion is one to one. For conversion in either direction the commutation conditions are sufficient but may not be necessary, since other methods not based on commutation might still be able to give an equivalent MRB.

MRB conversion allows taking advantage of the property tradeoffs between hierarchic and parallel MRB architectures. If speed is the primary consideration, a hardwired parallel architecture can be used. If the hardware is limited, a hardwired hierarchic architecture is a better choice. On serial computers, the hierarchy is both faster and requires less storage for filter coefficients than the parallel architecture. The conversions are easily extended to image signals if rectangular sampling is used. The conversions are also useful in the modeling of naturally occurring systems, such as parts of the nervous system [8].

In practice, many practical MRB's easily satisfy the commutation and stability conditions. MRB's in the hierarchic literature commonly use low-pass or Gaussian filters and have relatively prime factors. These MRB's always fulfill the commutation and stability conditions. Other MRB's normally use either low-pass or bandpass filters. MRB's with low-pass filters will satisfy the stability condition, since the filter cutoffs typically decrease with decreasing sampling rate. Parallel MRB's with bandpass filters do not satisfy the condition, but could be implemented with difference-pyramid type hierarchies [1], [2], [12], [14], [15].

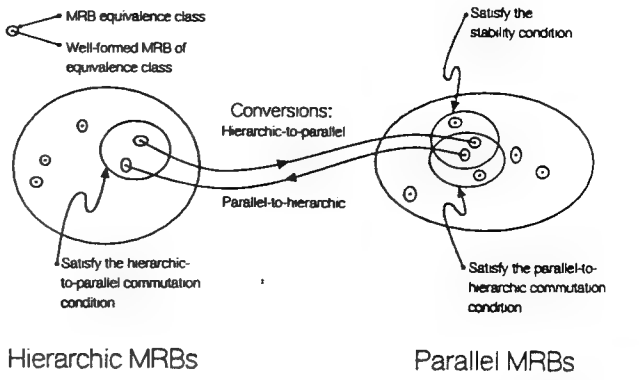


Fig. 5. Graphic summary of conditions for conversion. The spaces of all hierarchic and parallel MRB's are represented by the large ovals. Example MRB equivalence classes are indicated by the small ovals. The dot within a small oval represents the single well-formed MRB within that class.

References

[1] P. J. Burt, "The pyramid as a structure for efficient computation," in *Multiresolution Image Processing and Analysis*, A. Rosenfeld, Ed. Berlin: Springer-Verlag, 1984, pp. 6–35.

[2] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, no. 4, Apr. 1983.

[3] S. Chu and C. S. Burrus, "Multirate filter designs using comb filters," *IEEE Trans. Circuits Syst.*, vol. CAS-31, no. 11, pp. 913–924, Nov. 1984.

[4] S. Chu and C. S. Burrus, "Optimum FIR and IIR multistage multirate filter design," *Circuits, Syst., Signal Processing*, vol. 2, no. 3, pp. 361–386, July 1983.

[5] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[6] R. E. Crochiere, S. A. Webber, and J. L. Flanagan, "Digital coding of speech in subbands," *Bell Syst. Tech. J.*, vol. 55, no. 8, pp. 1069–1085, 1976.

[7] J. Kovačević and M. Vetterli, "Perfect reconstruction filter banks with rational sampling rates in one and two dimensions," *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 1199, pp. 1258–1268, 1989.

[8] B. Levitan and G. Buchsbaum, "A multirate filter bank perspective of early vision," *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 1199, pp. 1193–1202, 1989.

[9] B. Levitan and G. Buchsbaum, "Architecture-dependent properties of analysis multirate filter banks," in *Proc. ICASSP*, 1991, pp. 1805–1808.

[10] B. Levitan and G. Buchsbaum, "Conversions between parallel and hierarchic architecture analysis multirate filter banks," Tech. Rep. BE-VIS-191, Dep. Bioeng., Univ. Pennsylvania, 1991.

[11] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.

[12] L. O'Gorman and A. C. Sanderson, "A comparison of methods and computation for multiresolution low and bandpass transforms for image processing," *Comput. Graphics Image Processing*, vol. 37, pp. 386–401, 1987.

[13] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[14] S. Peleg, O. Federbusch, and R. Hummel, "Custom-made pyramids," in *Parallel Computer Vision*, L. Uhr, Ed. New York: Academic, pp. 125–146, 1987.

[15] A. Rosenfeld, Ed., *Multiresolution Image Processing and Analysis*. Berlin: Springer-Verlag, 1984.

[16] A. Tran and K. Liu, "An efficient pyramid image coding system," in *Proc. ICASSP*, 1987, pp. 744–747.

[17] P. P. Vaidyanathan, "Multirate digital filters, filter banks, polyphase networks, and applications: A tutorial," *Proc. IEEE*, vol. 78, no. 1, Jan. 1990.

[18] M. Vetterli, "Running FIR and IIR filtering using multirate filter banks," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, no. 5, pp. 730–738, May 1988.

[19] A. B. Watson, "The cortex transform: Rapid computation of simulated neural images," *Comput. Graphics Image Processing*, vol. 39, no. 3, Sept. 1987.

[20] J. W. Woods and S. D. O'Neil, "Subband coding of images," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 5, pp. 1278–1288, Oct. 1986.

JOURNAL OF VISUAL COMMUNICATION AND IMAGE REPRESENTATION
Vol. 4, No. 2, June, pp. 187–195, 1993

Appendix F

Conversion and Trade-offs between Scaled Gaussian Parallel and Hierarchic Analysis Multirate Filter Banks[1]

BENNETT LEVITAN AND GERSHON BUCHSBAUM

Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, 220 South 33rd St., Philadelphia, Pennsylvania 19104-6392

Received September 20, 1991; accepted July 22, 1992

Scaled Gaussian analysis multirate filter banks (MRBs) are analysis MRBs whose filters are Gaussians scaled differently in width and height. They are frequently employed in image processing and visual systems modeling. We define generalized scaled Gaussian analysis MRBs for both parallel and hierarchic architectures and derive closed-form equations for conversions between them. The MRBs have arbitrary rational number changes in sampling rate between successive outputs, and arbitrary vertical and horizontal scaling of the Gaussian filters for each output. We calculate the number of multiplications, number of additions, and total number of filter coefficients to compare the parallel and hierarchic architectures as implemented with direct form and time-varying filters. In all cases, the parallel MRB requires considerably more multiplications, additions, and filter coefficients than the equivalent hierarchic MRB. However, the relative differences are far less severe in the time-varying case. We also derive a useful approximation for the discrete time Fourier transform of a Gaussian.

INTRODUCTION

To detect and manipulate image features, many image processing systems operate on several spatial scales. For example, image recognition systems apply the same detection algorithm to small and large versions of the same objects [1–4]. It is often beneficial to distinguish between spatial scales. The technique of subband coding, for instance, operates efficiently by differentially allocating processing and storage resources to the scales [1, 5–9]. In many cases, multispatial scale processing is performed with analysis multirate filter banks (MRBs) [1, 5, 8, 10–15]. Analysis MRBs consist of a set of filters and resamplers that produce several different sampling rate versions of an image. They attain great computational savings by manipulating large spatial scale features in low sampling rate outputs. Analysis MRBs can be implemented with parallel, hierarchic or other architectures. In the parallel architecture, each output is produced by filters and resamplers acting on the input image. In the hierarchic architecture, the nth output is produced by filters and resamplers acting on the $n-1$st output.

An analysis MRB often used in image processing is the scaled Gaussian analysis MRB [5, 10–14]. These are MRBs all of whose filter impulse responses are Gaussians scaled differently in width and height. Reasons that scaled Gaussian MRBs are popular include: (i) Two-dimensional Gaussians are separable and hence easily implemented. (ii) As shown below, filters of the MRB equivalent to a given scaled Gaussian MRB of the other architecture are always Gaussian. (iii) As also shown below, provided the Gaussians of a parallel MRB increase sufficiently in width between successive outputs, the equivalent hierarchic filters are real. With arbitrary filters, the hierarchic equivalent filters may be IIR, complex, or unstable. (iv) The $\nabla^2 G$ operator commonly used in image processing can be well matched by taking the difference between appropriately Gaussian-filtered images [16]. (v) Gaussians are very useful in modeling the receptive fields of cells in mammalian visual systems [17–21]. Parts of the visual systems can be modeled as an MRB with Gaussian or related filters [13, 14, 21, 22].

In this paper, we derive conversions between scaled Gaussian hierarchic and parallel MRBs and calculate some of their properties. Conversions and properties have been derived for some MRBs, but not for the general case [5, 10–12, 15, 23, 24]. The MRBs that have been examined downsample by the same factor between outputs and change the sampling rate only by integer factors. Conversions and properties for more general cases were derived in [25, 26]. In this paper, Gaussians have arbitrary vertical and horizontal scaling for each output, and the sampling rate may change by a different rational num-

[1] This work supported by Grant AFOSR-91-00-82 from the Air Force Office of Scientific Research. B. Levitan was also supported by a National Institutes of Health fellowship under Grant 5-T32-GM07170.

1047-3203/93 $5.00

ber between each output. We compute number of multiplications, number of additions and required filter storage for both direct form and time-varying filter implementations.

THEORY

Conversion between Architectures

In this section, we derive the conversions for the scaled Gaussian case. For simplicity, the derivations below will be for one-dimensional signals. Generalization to two-dimensional signals with separable filters, like Gaussians, is straightforward, and we give results and properties for both types of signals. To distinguish between parallel and hierarchic MRB variables, all parallel MRB variables are in boldface type. In the parallel MRB, each of the N outputs $\mathbf{y}_n(x_n)$ is computed from the original input $y_0(x_0)$ in three steps: (i) sampling rate expansion by factor $\mathbf{L}_n$, (ii) filtering by LTI parallel filter $\mathbf{h}_n(\mathbf{u}_n)$, and (iii) sampling rate compression by factor $\mathbf{M}_n$ (Fig. 1a). In the hierarchic MRB, outputs $y_1(x_1)$ through $y_N(x_N)$ are produced successively in N stages. The nth stage computes y_n from y_{n-1} in three steps: (i) sampling rate expansion by factor L_n, (ii) filtering by LTI parallel filter $h_n(u_n)$, and (iii) sampling rate compression by factor M_n (Fig. 1b).

Parallel to Hierarchic Conversion. To incorporate scaled filters into a parallel MRB, the filter impulse responses are defined as scaled, sampled versions of prototype function $f_c(t)$ where t is a continuous variable [26],

$$\mathbf{h}_n(\mathbf{u}_n) = \mathbf{A}_n f_c(\mathbf{u}_n/\mathbf{B}_n) \quad n = 1, \ldots, N, \tag{1}$$

where $\mathbf{A}_n$ and $\mathbf{B}_n$ are positive, real-valued coefficients that allow arbitrary vertical and horizontal scaling of f_c; f_c must have a continuous argument since the $\mathbf{B}_n$ are arbitrary. From [26], provided basic conditions for conversion are met, if we define discrete function $f_n(x) = f_c(x/\mathbf{B}_n)$ and $F_n(\omega)$ as its discrete time Fourier transform (DTFT), the hierarchic filters are the inverse DTFT of

$$H_n(\omega) = \frac{\mathbf{A}_n \mathbf{F}_n(\omega C_n/\mathbf{M}_{n-1})}{\mathbf{A}_{n-1}\mathbf{F}_{n-1}(\omega L_n/\mathbf{M}_{n-1})}, \tag{2}$$

where $C_n = \mathbf{L}_{n-1}\,\mathbf{M}_{n-1}/[\gcd(\mathbf{L}_n, \mathbf{L}_{n-1})\,\gcd(\mathbf{M}_n, \mathbf{M}_{n-1})]$ and $\gcd(a, b)$ is the greatest common divisor of a and b. L_n in (2) is the equivalent hierarchic upsampling factor for stage n. It and the equivalent downsampling factors are given by

$$L_1 = \mathbf{L}_1, \quad L_n = C_n \mathbf{L}_n/\mathbf{L}_{n-1}, \quad n = 2, \ldots, N, \tag{3}$$

$$M_1 = \mathbf{M}_1, \quad M_n = C_n \mathbf{M}_n/\mathbf{M}_{n-1}, \quad n = 2, \ldots, N. \tag{4}$$

For scaled Gaussian filters, $f_c(t)$ is defined as $f_c(t) = \exp(-t^2/\sigma^2)$. Substituting f_c into (1) gives the parallel filters $\mathbf{h}_n(\mathbf{u}_n) = A_n \exp(\mathbf{u}_n^2/\sigma^2\mathbf{B}_n^2)$, $n = 1, \ldots, N$. To solve for the equivalent hierarchic filters, we note that the DTFT of $f_n(x) = \exp(-x^2/\sigma^2\mathbf{B}_n^2)$ is

$$F_n(\omega) = \sqrt{\pi}\,\sigma\,\mathbf{B}_n \sum_{m=-\infty}^{\infty} \exp\left(\frac{-\sigma^2\mathbf{B}_n^2\,(\omega - 2\pi m)^2}{4}\right) \tag{5}$$

In the Appendix, we show that (5) can be excellently approximated in the interval $-\pi \le \omega \le \pi$ by $F_n(\omega) \approx \sqrt{\pi}\,\sigma\,\mathbf{B}_n \exp(-\sigma^2\mathbf{B}_n^2\omega^2/4)$ for $\sigma\mathbf{B}_n \ge 1.8981$ to within 0.1%. Substitution into (2) gives hierarchic filters

$$H_n(\omega) \approx \frac{\mathbf{A}_n\mathbf{B}_n}{\mathbf{A}_{n-1}\mathbf{B}_{n-1}} \exp\left(-\frac{\sigma^2(\mathbf{B}_n^2C_n^2 - \mathbf{B}_{n-1}^2L_n^2)\,\omega^2}{4\mathbf{M}_{n-1}^2}\right) \quad n = 2, \ldots, N. \tag{6}$$

With the DTFT approximation, we can consider (6) as a constant multiplied by the approximated DTFT of a Gaussian with variance $\sigma^2(\mathbf{B}_n^2C_n^2 - \mathbf{B}_{n-1}^2L_n^2)/\mathbf{M}_{n-1}^2$. The inverse transform is

$$h_n(u_n) \approx \frac{\mathbf{A}_n\mathbf{B}_n\mathbf{M}_{n-1}}{\mathbf{A}_{n-1}\mathbf{B}_{n-1}\sqrt{\pi}\,\sigma(\mathbf{B}_n^2C_n^2 - \mathbf{B}_{n-1}^2L_n^2)^{1/2}} \exp\left(-\frac{\mathbf{M}_{n-1}^2u_n^2}{\sigma^2(\mathbf{B}_n^2C_n^2 - \mathbf{B}_{n-1}^2L_n^2)}\right) \quad n = 2, \ldots, N. \tag{7}$$

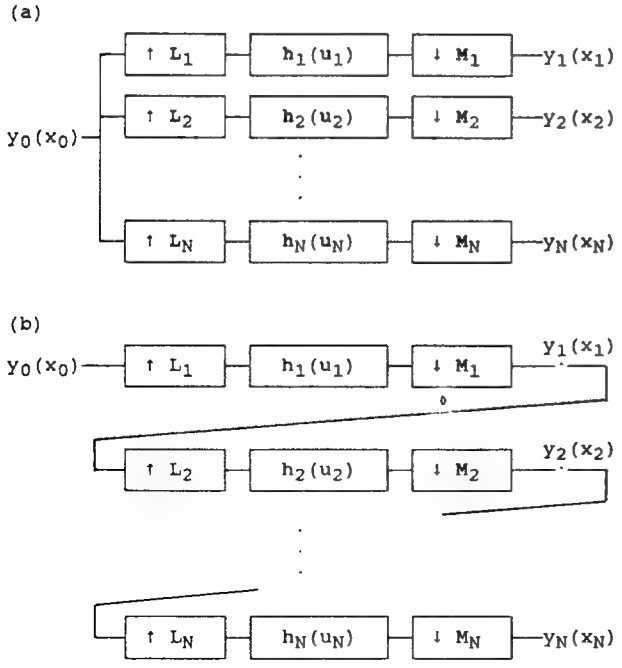


FIG. 1. Multirate filter bank architectures: (a) parallel MRB and (b) hierarchic MRB.

Equation (7) gives the $n \geq 2$ filters for the hierarchic MRB equivalent to a given parallel MRB with scaled Gaussian filters. For $n = 1$, $h_1(u_1) = \mathbf{h}_1(u_1)$.

Because $H_n(\omega)$ has no poles, the $h_n(u_n)$ are FIR and stable. The hierarchic filters are Gaussian, since they result from the convolution of two Gaussians. Examples of filters from equivalent parallel and hierarchic MRBs are shown later in Fig. 3. If $\mathbf{B}_n \leq \mathbf{B}_{n-1}\mathbf{L}_n/\mathbf{L}_{n-1}$, then $\mathbf{B}_n^2 C_n^2 - \mathbf{B}_{n-1}^2 L_n^2$ in the denominator of (7) is negative or zero, and the equivalent hierarchic filters are undefined or complex. In practice, this situation can be avoided by proper choice of $\mathbf{B}_n$ or reordering the outputs.

Hierarchic to Parallel Conversion. As in the parallel case, to incorporate scaled filters into a hierarchic MRB, the filter impulse responses are defined as scaled versions of a continuous prototype function $f_c(t)$;

$$h_n(u_n) = A_n f_c(u_n/B_n), \quad n = 1, \ldots, N, \tag{8}$$

where again A_n and B_n are positive, real-valued coefficients that allow arbitrary vertical and horizontal scaling of f_c. From [26], defining discrete function $f_n(x) = f_c(x/B_n)$ and $F_n(\omega)$ as its DTFT, $H_n(\omega) = A_n F_n(\omega)$, and the equivalent parallel filter is

$$\mathbf{H}_n(\omega) = \prod_{k=1}^{n} A_k \prod_{j=1}^{n} F_n(\omega W_{j,n}), \tag{9}$$

where

$$W_{1,1} = 1, \quad W_{j,n} = \prod_{i=1}^{j-1} M_i \prod_{i=j+1}^{n} L_i, \tag{10}$$

and M_i and L_j are relatively prime, $i = 1, \ldots, N$, $j = i + 1, \ldots, N$. The constraints on M_i and L_j are sufficient to satisfy the basic conditions for hierarchic to parallel conversion and allows us to easily write the closed-form solutions given in (9) [26]. For scaled Gaussian filters, we set $f_c(t) = e^{-t^2/\sigma^2}$. Substituting this into (8) gives $h_n(u_n) = A_n\, e^{-u_n^2/\sigma^2 B_n^2}$, $n = 1, \ldots, N$. From the Appendix, the DTFT of the Gaussian $f_n(x)$ for $B_n\sigma \geq 1.8981$ is $F_n(\omega) \approx \sqrt{\pi}\ \sigma\ B_n\ e^{-\sigma^2 B_n^2 \omega^2/4}$. Substitution into (9) gives

$$\mathbf{H}_n(\omega) \approx \sigma^n B_n^{n\pi^{n-1}} \left(\prod_{k=1}^{n} A_k\right) \exp\left(-\frac{\sigma^2 B_n^2 \omega^2}{4} \sum_{j=1}^{n} W_{j,n}^2\right). \tag{11}$$

Using the approximation, (11) can be regarded as a constant multiplied by the DTFT of Gaussian with variance $\sigma^2 B_n^2 \sum_{j=1}^{n} W_{j,n}^2$. Its inverse transform yields the parallel filters and completes the conversion:

$$\mathbf{h}_n(\mathbf{u}_n) \approx \sigma^{n-1} B_n^{n-1\pi^{(n-1)/2}} \left(\sum_{j=1}^{n} W_{j,n}^2\right)^{-1/2} \left(\prod_{k=1}^{n} A_k\right) \exp\left(-\mathbf{u}_n^2 \Big/ \left[\sigma^2 B_n^2 \sum_{j=1}^{n} W_{j,n}^2\right]\right). \tag{12}$$

Equation (12) gives the filters for the parallel MRB equivalent to a given scaled Gaussian hierarchic MRB. The equivalent resampling factors derived in [26] are

$$\mathbf{L}_n = \prod_{i=1}^{n} L_i / \gcd\left(\prod_{i=1}^{n} L_i, \prod_{i=1}^{n} M_i\right), \qquad \mathbf{M}_n = \prod_{i=1}^{n} M_i / \gcd\left(\prod_{i=1}^{n} L_i, \prod_{i=1}^{n} M_i\right). \tag{13}$$

Example Conversion. We demonstrate the conversions by implementing a four-output scaled Gaussian parallel MRB and its equivalent hierarchic MRB as derived above. Table 1 lists the factors and filter lengths used. $\mathbf{A}_n$ were chosen so each Gaussian has volume equal to $\mathbf{L}_n$. These values maintain the DC value of the input signal in each output. σ and $\mathbf{B}_n$ were chosen to make $\mathbf{H}_n(\omega)$ negligible beyond $\pi/\max(\mathbf{M}_n, \mathbf{L}_n)$, the largest frequency that can be passed without aliasing or imaging [27]. $\mathbf{M}_n = 3^n$ and $\mathbf{L}_n = 2^n$ for all four outputs. Parallel MRB filter lengths were selected so the equivalent hierarchic filters included greater than 99.9% of the area they would occupy if infinitely extended.

Figure 2 shows the input image and outputs of the MRBs. Visually, the corresponding outputs are identical. All root-mean-square error differences between corresponding outputs are very small, and are caused mostly by edge effects. However, successive outputs show an increasing rms difference. We attribute this trend to a problem fundamental to hierarchic processors: Since the images and filters are quantized and the filters are of finite length, each output has a small amount of quantization error and aliasing error. In a hierarchic processor, these errors propagate from level to level and accumulate. While parallel processors also have quantization errors and aliasing, the errors do not propagate. For this reason,

TABLE 1
Factors and Filter Lengths Used in the Example of Scaled Gaussian MRBs ($\sigma = 2$)

n	$\mathbf{A}_n$	$\mathbf{B}_n$	$\mathbf{L}_n$	$\mathbf{M}_n$	L_n	M_n	$\mathbf{t}_n$	t_n
1	0.06577	2.2	2	3	2	3	21	21
2	0.01989	8.0	4	9	2	3	101	21
3	0.01052	22.0	8	27	2	3	345	17
4	0.00513	63.0	16	81	2	3	540	17

FIG. 2. Input and outputs of equivalent parallel and hierarchic MRBs: (a) Input image (256 × 256 pixels), (b) outputs of parallel MRB (171 × 171, 114 × 114, 76 × 76, 51 × 51 pixels), and (c) outputs of equivalent hierarchic MRB.

successive images in the hierarchic and parallel MRBs may increasingly differ.

Figure 3 shows the filters used to make these outputs. Each filter ($\mathbf{h}_n$ or h_n) is plotted against its own spatial axis ($\mathbf{u}_n$ or u_n). While h_1 and h_2 are similar as are h_3 and h_4, there is in general no reason for the hierarchic filters to be similar. h_1 is identical to $\mathbf{h}_1$. The other h_n differ from the $\mathbf{h}_n$ in several respects: (i) Their peak amplitudes are larger; (ii) Their energy is much more tightly compressed about the origin; (iii) While the parallel filters are successively more low pass, the hierarchic filters do not become more or less low pass in any order. These differences are due to the incremental nature of the hierarchy.

Properties of Architectures

The computational complexity of a system is the number of multiplications and additions it requires. In this section, we calculate the computational complexity and total filter storage required for the parallel and hierarchic MRBs as computed by a serial (single instruction stream, single data stream) machine. We give results for both one-dimensional MRBs and two-dimensional MRBs using separable filters. As in many applications, the "parallel" architecture, in which each output is computed directly from the input, is actually implemented on a serial computer; only one output is calculated at a time. The

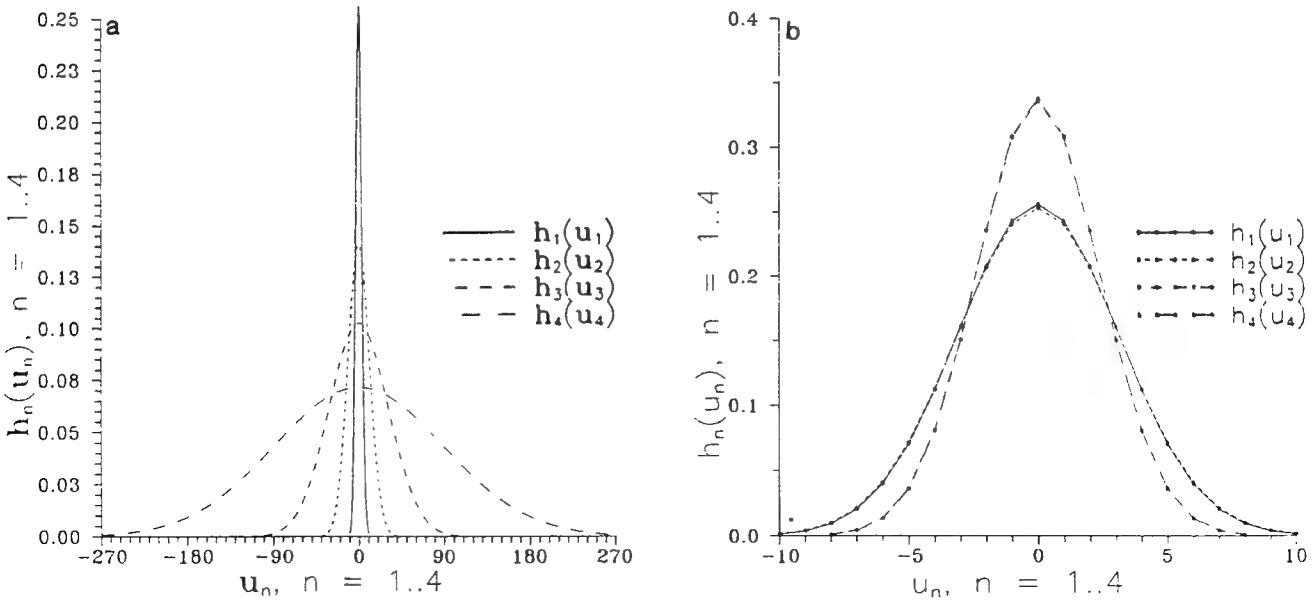


FIG. 3. Filters from equivalent scaled Gaussian MRBs. Each impulse response is plotted against its own spatial axis: (a) parallel filters and (b) hierarchic filters.

complexity for the parallel architecture is thus the sum of the complexities for each output. In a true parallel implementation, in which the N outputs are computed simultaneously, the complexity is the maximum of the complexities for each output.

The properties are based on MRB filter lengths. Since parallel filter $\mathbf{h}_n$ is the convolution of sampling rate-expanded versions of hierarchic filters h_1 through h_n (9) [26],

$$\mathbf{t}_n = 1 + \sum_{j=1}^{n} W_{j,n}(t_j - 1). \tag{14}$$

where $\mathbf{t}_n$ is the number of coefficients in $\mathbf{h}_n$, t_n is the number of coefficients in h_n, and the $W_{j,n}$ are as defined in (10). Since the Gaussian filters are FIR, we can use (10) to solve (14) for finite t_n, giving

$$t_n = \frac{\mathbf{t}_n - L_n(\mathbf{t}_{n-1} - 1) - 1}{W_{n,n}} + 1. \tag{15}$$

Parallel MRBs that can be converted to equivalent hierarchic MRBs yield integer t_n in (15). For simplicity in the derivations, we consider filter impulse responses of odd length and later show the substitution necessary for even-length filters.

We define S_n as the number of samples in signal y_n. From [27], in the parallel MRB

$$S_n = \lceil S_0 \mathbf{L}_n / \mathbf{M}_n \rceil, \tag{16}$$

and in the hierarchic MRB,

$$\begin{aligned} S_n &= \lceil S_{n-1} L_{n-1} / M_{n-1} \rceil = \lceil \lceil S_{n-2} L_{n-2} / M_{n-2} \rceil L_{n-1} / M_{n-1} \rceil \\ &= \cdots = \lceil \cdots \lceil \lceil S_0 L_1 / M_1 \rceil L_2 / M_2 \rceil \cdots L_{n-1} / M_{n-1} \rceil, \end{aligned} \tag{17}$$

where $\lceil x \rceil$ indicates the smallest integer greater than or equal to x. Though $\mathbf{L}_n/\mathbf{M}_n = L_1 L_2 \cdots L_n / M_1 M_2 \cdots M_n$ (Eq. (13)), Eqs. (16) and (17) do not always give the same results. However, the differences are always small or zero for typical resampling factors. To compare properties and results in different architectures, we truncate the outputs at the smaller of the two sizes those given by (16). The two-dimensional nth outputs contains S_n^2 samples.

Direct Form Filters. To calculate the computational complexity of the MRB, we total the number of multiplications and additions in all the stages. In the nth stage of a parallel MRB, y_0 is first expanded by $\mathbf{L}_n$ and filtered by $\mathbf{h}_n$. With direct form filters, the stage performs a total of $S_0 \mathbf{L}_n \mathbf{t}_n$ multiplications, or $\mathbf{L}_n \mathbf{t}_n$ multiplications per input sample (with appropriate padding of the input signal at the edges). Taking advantage of symmetry in the odd-length Gaussian filters, there are only $\mathbf{L}_n[(\mathbf{t}_n - 1)/2 + 1]$ multiplications per input sample. Thus, the total number of multiplications per input sample in the parallel MRB is

$$m_{\text{df,p}} = \sum_{k=1}^{N} \mathbf{L}_k[(\mathbf{t}_k - 1)/2 + 1]. \tag{18}$$

In this notation, "df" or "tv" stand for direct form or time-varying filters, and "p" or "h" stand for the parallel or hierarchic architectures. The number of additions in a filter is one less than the number of filter coefficients, thus the total number of additions is

$$a_{\text{df,p}} = \sum_{k=1}^{N} \mathbf{L}_k(\mathbf{t}_k - 1). \tag{19}$$

As discussed below, the number of filter coefficients depends only on the architecture (parallel vs. hierarchic) and not on whether the filters are direct form or time-varying. The total number of coefficients is the sum of the number of independent coefficients in all filters;

$$c_{\text{p}} = \sum_{k=1}^{N} \left[\frac{(\mathbf{t}_k - 1)}{2} + 1 \right], \tag{20}$$

where again we have taken advantage of filter symmetry.

For two-dimensional signals, the results are similar. Since the filters are separable, the $S_0 \mathbf{L}_n[(t_n - 1)/2 + 1]$ multiplications for the one-dimensional signal are repeated S_0 times along one axis and S_0 times along the other. This gives $2S_0^2 \mathbf{L}_n[(\mathbf{t}_n - 1)/2 + 1]$ multiplications for the nth stage, or $2\mathbf{L}_n[(\mathbf{t}_n - 1)/2 + 1]$ multiplications per input sample. Thus, for two-dimensional signals, the number of multiplications and additions (Eqs. (18) and (19) are doubled. However, since the same filter is used along each axis, the number of filter coefficients (20) remains the same.

For the one-dimensional, hierarchic MRB, the nth stage performs $S_{n-1} L_n[(t_n - 1)/2 + 1]$ multiplications. To normalize by the number of samples in y_0, we approximate (16) with $S_n/S_0 = \mathbf{L}_n/\mathbf{M}_n$. This approximation differs from (16) by at most $1/S_0$ and is excellent, since $S_0 \gg 1$. For the first stage, dividing out S_0 gives $L_1[(t_1 - 1)/2 + 1]$ multiplications per input sample. For all other stages, there are $(\mathbf{L}_{n-1}/\mathbf{M}_{n-1}) L_n[(t_n - 1)/2 + 1]$ multiplications per input sample. Thus, the total number of multiplications and additions per input sample in the one-dimensional hierarchic MRB is

$$m_{\text{df,h}} = L_1 \left[\frac{t_1 - 1}{2} + 1 \right] + \sum_{k=2}^{N} \frac{\mathbf{L}_{k-1}}{\mathbf{M}_{k-1}} L_k \left[\frac{t_k - 1}{2} + 1 \right] \tag{21}$$

TABLE 2
Properties for Direct form the Time-Varying Implementations of Equivalent Parallel and Hierarchic MRBs ($N = 4$): Two Examples Are Shown; the Signal Is One-Dimensional

					Direct form		Time-varying	
L	M	t	Architecture	No. of Coefficients	No. of mults/sample	No. of adds/sample	No. of mults/sample	No. of adds/sample
2	3	9	Parallel	364	4,886	9,712	19.65	18.05
			Hierarchic	20	24	39	7.22	5.62
3	4	9	Parallel	884	61,080	121,920	15.93	13.88
			Hierarchic	20	41	66	6.15	4.10

$$a_{\text{df,h}} = L_1(t_k - 1) + \sum_{k=2}^{N} \frac{\mathbf{L}_{k-1}}{\mathbf{M}_{k-1}} L_k(t_k - 1). \quad (22)$$

The total number of independent filter coefficients is

$$c_{\text{h}} = \sum_{k=1}^{N} \left[\frac{(t_k - 1)}{2} + 1\right]. \quad (23)$$

For even-length filters, (19) and (22) remain unchanged, and Eqs. (18), (20), (21), and (23) can be used if $(\mathbf{t}_k - 1)/2 + 1]$ is replaced by $\mathbf{t}_k/2$ and $(t_k - 1)/2 + 1]$ is replaced by $t_k/2$. For two-dimensional signals, the results are slightly more complex than in the parallel case. The nth filter performs $2S_{n-1}^2 L_n[(t_n - 1)/2 + 1]$ multiplications, or $2(\mathbf{L}_{n-1}/\mathbf{M}_{n-1})^2 L_n[(t_n - 1)/2 + 1]$ multiplications per input sample. The results for the entire two-dimensional, hierarchic MRB are

$$m_{\text{df,h,2D}} = 2L_1\left[\frac{t_1 - 1}{2} + 1\right] + 2\sum_{k=2}^{N}\left(\frac{\mathbf{L}_{k-1}}{\mathbf{M}_{k-1}}\right)^2 L_k\left[\frac{t_k - 1}{2} + 1\right] \quad (24)$$

$$a_{\text{df,h,2D}} = 2L_1(t_k - 1) + 2\sum_{k=2}^{N}\left(\frac{\mathbf{L}_{k-1}}{\mathbf{M}_{k-1}}\right)^2 L_k(t_k - 1) \quad (25)$$

where the "2D" indicates the result is for two-dimensional signals. As in the parallel case, the number of filter coefficients is the same for one and two-dimensional signals.

Table 2 shows these properties for two example MRBs. In all cases, the MRBs are based on a hierarchic MRB having $t_n = t$, $L_n = L$, and $M_n = M$ for $n = 1, \ldots, N$, the most common type of hierarchic MRB used. As shown in the table, $m_{\text{df,p}}$ and $a_{\text{df,p}}$ exceed $m_{\text{df,h}}$ and $a_{\text{df,h}}$ by several orders of magnitude, because the parallel filters are so much larger than the equivalent hierarchic filters. The ratios of $c_{\text{df,p}}$ to $c_{\text{df,h}}$ also reflect this difference. The parallel values are larger because the parallel MRB does in single steps the same filtering that the hierarchic MRB does incrementally. Table 3 shows that the relative values for the two-dimensional MRB properties are similar to those for the one-dimensional case.

Time-Varying Filters. Direct form filters can be useful since they are easy to design and require little overhead computation. However, they are highly inefficient for multirate filtering. Because the filtering in each stage is performed after the sampling rate expander, where the sampling rate is highest: (i) the $\mathbf{L}_n - 1$ zeros inserted between input samples by the expander (in a parallel-type stage) are multiplied by the filter weights, even though the zeros do not influence the outputs, and (ii) $\mathbf{M}_n - 1$ of

TABLE 3
Properties for Direct form and Time-Varying Implementations of Equivalent Parallel and Hierarchic MRBs ($N = 4$): Two Examples Are Shown; the Signal Is Two-Dimensional

					Direct form		Time-varying	
L	M	t	Architecture	No. of Coefficients	No. of mults/sample	No. of adds/sample	No. of mults/sample	No. of adds/sample
2	3	9	Parallel	364	9,772	19,424	39.31	36.10
			Hierarchic	20	35	55	10.38	5.92
3	4	9	Parallel	884	122,160	243,840	31.85	27.75
			Hierarchic	20	62	99	9.26	4.61

every $\mathbf{M}_n$ outputs $\mathbf{h}_n$ calculates are ignored by the compressor.

Time-varying filters are much more efficient [8, 27]. At the cost of some overhead computation, they ignore the inserted zeros and only calculate the samples passed by the compressor. They achieve these savings by cycling among $\mathbf{L}_n$ different impulse functions, each composed of different samples from the direct form filter $\mathbf{h}_n$. These impulse responses are either $\lceil \mathbf{t}_n/\mathbf{L}_n \rceil$ or $\lfloor \mathbf{t}_n/\mathbf{L}_n \rfloor$ samples in length, where $\lfloor x \rfloor$ indicates the largest integer less than or equal to x. Their average size is $\mathbf{t}_n/\mathbf{L}_n$, which we use below as filter length. Thus, the number of multiplications in the nth stage is $(S_0\mathbf{L}_n/\mathbf{M}_n)(\mathbf{t}_n/\mathbf{L}_n)$, or $\mathbf{t}_n/\mathbf{M}_n$ multiplications per input sample, where $S_0\mathbf{L}_n/\mathbf{M}_n$ is the number of output samples computed. Because the h_n is asymmetrically broken into $\mathbf{L}_n$ functions, it is not possible to take advantage of its symmetry. The number of additions for the nth stage is $(S_0\mathbf{L}_n/\mathbf{M}_n)(\mathbf{t}_n/\mathbf{L}_n - 1)$, or $(\mathbf{L}_n/\mathbf{M}_n)(\mathbf{t}_n/\mathbf{L}_n - 1)$ additions per input sample. For the time-varying, one-dimensional, parallel MRB, the total number of multiplications and additions per input sample is

$$m_{\text{tv,p}} = \sum_{k=1}^{N} \mathbf{t}_k/\mathbf{M}_k \tag{26}$$

$$a_{\text{tv,p}} = \sum_{k=1}^{N} \frac{\mathbf{L}_k}{\mathbf{M}_k} (\mathbf{t}_k/\mathbf{L}_k - 1). \tag{27}$$

For two-dimensional parallel MRBs, the number of multiplications and number of additions are double the one-dimensional values. Since the $\mathbf{L}_n$ impulse functions are composed of samples from the direct form filter, for both the one- and two-dimensional time-varying filter, parallel cases, the total number of independent filter coefficients stored is the same as for the direct form, parallel case. For the time-varying, one-dimensional hierarchic MRB, the nth stage performs $(S_{n-1}L_n/M_n)(t_n/L_n)$ multiplications and $(S_{n-1}L_n/M_n)(t_n/L_n - 1)$ additions. Thus, the number of multiplications and additions per input sample are

$$m_{\text{tv,h}} = t_1/M_1 + \sum_{k=2}^{N} (\mathbf{L}_{k-1}/\mathbf{M}_{k-1})(t_k/M_k) \tag{28}$$

$$a_{\text{tv,h}} = (L_1/M_1)(t_1/L_1 - 1) + \sum_{k=2}^{N} (\mathbf{L}_{k-1}/\mathbf{M}_{k-1})(L_k/M_k)(t_k/L_k - 1). \tag{29}$$

For two-dimensional signals, the results are

$$m_{\text{tv,h,2D}} = 2t_1/M_1 + 2\sum_{k=2}^{N} \left(\frac{\mathbf{L}_{k-1}}{\mathbf{M}_{k-1}}\right)^2 t_k/M_k \tag{30}$$

$$a_{\text{tv,h,2D}} = 2(L_1/M_1)(t_1/L_1 - 1) + 2\sum_{k=2}^{N} \left(\frac{\mathbf{L}_{k-1}}{\mathbf{M}_{k-1}}\right)^2 (L_k/M_k)(t_k/L_k - 1). \tag{31}$$

The total number of filter coefficients stored for both the one- and two-dimensional time-varying, hierarchic cases is the same as for the direct form, hierarchic case.

Table 2 and 3 show these properties for the same MRBs as used in the direct form case. As shown in the tables, $m_{\text{tv,p}}$ and $a_{\text{tv,p}}$ exceed $m_{\text{tv,h}}$ and $a_{\text{tv,h}}$, but by a much smaller factor than in the direct form case. While the parallel filters are still much larger than the equivalent hierarchic filters, the efficient time-varying implementation lessens the significance of this difference.

CONCLUSION

This paper gives closed-form conversions between parallel and hierarchic scaled Gaussian MRBs and calculates properties of these architectures. We have shown that the hierarchic architecture always requires fewer multiplications, additions, and coefficients than the equivalent parallel architecture. The hierarchy's advantages are much smaller for time-varying filters than for direct form filters. However, hierarchic MRBs can suffer from the propagation and accumulation of quantization and aliasing error in successive stages. Additionally, in hardwired implementations, while the hierarchic MRB requires far fewer hardwired multiplications, the parallel MRB operates more rapidly [25]. Thus, despite its greater complexity, there are applications for which the parallel architecture may be more useful. Conversion allows taking advantage of these and other differences between the architectures.

APPENDIX

Approximation of Discrete Time Fourier Transform of a Gaussian

The Discrete Time Fourier Transform of the Gaussian $\exp(-x^2/\sigma^2)$ is

$$\sqrt{\pi}\,\sigma \sum_{m=-\infty}^{\infty} \exp\left(\frac{-\sigma^2(\omega - 2\pi m)^2}{4}\right). \tag{32}$$

Equation (32) is the sum of an infinite number of Gaussians displaced by $\pm 2\pi$. We are interested in the minimum σ such that the predominant contribution to (32) in the region $R = (-\pi \le \omega \le \pi)$ is the $n = 0$ Gaussian. To simplify the derivation, we drop the coefficient $\sqrt{\pi}\,\sigma$.

Let V_1 be the area under the center Gaussian in R

$$V_1 = \int_{-\pi}^{\pi} \exp(-\sigma^2\omega^2/4)\, d\omega = 2\sqrt{\pi}\, \mathrm{erf}(\pi\sigma/2)/\sigma \quad (33)$$

where erf() is the error function and V_2 be the area in R due to the other Gaussians

$$V_2 = \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \int_{-\pi}^{\pi} \exp\left(\frac{-\sigma^2(\omega - 2\pi n)^2}{4}\right) d\omega. \quad (34)$$

We define $V_2 = V_2^+ + V_2^-$ where V_2^+ is the sum in (34) over $n = 1$ to ∞, and V_2^- is the sum over $n = -1$ to $-\infty$. Due to symmetry, $V_2 = 2 \cdot V_2^+$. The problem can be restated as finding the minimum σ such that

$$(V_1 + V_2)/V_1 = (V_1 + 2 \cdot V_2^+)/V_1 = 1 + 2 \cdot V_2^+/V_1 = 1 + \delta, \quad 0 < \delta \ll 1. \quad (35)$$

For $n > 0$, the exponents in (34) are maximized in R at $\omega = \pi$ and $-\pi$. Thus, an upper bound for V_2^+ is

$$V_2^+ < 2\pi \sum_{n=1}^{\infty} \exp\left(\frac{-\sigma^2(\pi - 2\pi n)^2}{4}\right) = 2\pi \exp(-\sigma^2\pi^2/4) \sum_{n=1}^{\infty} (\exp(\sigma^2\pi^2))^{n(1-n)}. \quad (36)$$

Since σ and π are both positive, $\exp(\sigma^2\pi^2) > 1$. Noting that $\sum_{n=1}^{\infty} x^{n(1-n)} = 1 + \sum_{n=2}^{\infty} x^{n(1-n)} < 1 + \sum_{n=2}^{\infty} x^{-n}$, since $x^{-n} > x^{n(1-n)}$ for $x > 1$ and $n > 2$, substitution into (36) gives

$$V_2^+ < 2\pi \exp(-\sigma^2\pi^2/4) \left[1 + \sum_{n=2}^{\infty} (\exp(\sigma^2\pi^2))^{-n}\right]. \quad (37)$$

Applying the identity $\sum_{n=0}^{\infty} x^n = 1/(1 - x)$ for $|x| < 1$ to (37) gives

$$V_2^+ < 2\pi \exp(-\sigma^2\pi^2/4) \left[\frac{1}{1 - \exp(-\sigma^2\pi^2)} - \exp(-\sigma^2\pi^2)\right]. \quad (38)$$

Finally, substituting for V_1 and V_2^+ in (35) and simplifying gives

$$\frac{2\sigma\sqrt{\pi}}{\mathrm{erf}(\sigma\pi/2)} \left[\frac{1}{1 - \exp(-\sigma^2\pi^2)} - \exp(-\sigma^2\pi^2)\right] \leq \delta. \quad (39)$$

For $\delta = 0.001$, σ must be greater than approximately 1.8981. Thus, the approximation to (5) is excellent when $\sigma B_n \geq 1.8981$. By way of example, when $\sigma = 1.8981$, $V_1 = 3.735$. The contribution of the $n = 1$ Gaussians to V_2 is 4.634×10^{-5}. For $n > 1$, the contributions are even more negligible. This approximation is quite useful, since a Gaussian with standard deviation 1.8981 drops to below 0.001 its peak value only five samples from the origin.

REFERENCES

1. A. Rosenfeld (Ed.), *Multiresolution Image Processing and Analysis*, Springer-Verlag, Berlin, 1984.
2. P. J. Burt, Smart sensing within a pyramid vision machine, *IEEE Proc.* **76**(8), 1988, 1006–1014.
3. P. Saint-Marc, J. Chen, and G. Medioni, Adaptive smoothing: A general tool for early vision, *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 1991, 514–529.
4. W. Hoff and N. Ahuja, Surfaces from stereo: Integrating feature machine, disparity estimation, and contour detection, *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 1989, 121–136.
5. P. J. Burt and E. H. Adelson, The Laplacian pyramid as a compact image code, *IEEE Trans. Comm.* **COM-31**(4), 1983, 532–540.
6. R. E. Crochiere, S. A. Webber, and J. L. Flanagan, Digital coding of speech in sub-bands, *Bell Systems Tech. J.* **55**(8), 1976, 1069–1085.
7. S. Peleg, O. Federbusch, and R. Hummel, Custom-made pyramids, in *Parallel Computer Vision* (L. Uhr, Ed.), pp. 125–146, Academic Press, San Diego, 1987.
8. P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice Hall, Englewood Cliffs, NJ, 1993.
9. J. W. Woods and S. D. O'Neil, Subband coding of images, *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-34**(5), 1986, 1278–1288.
10. P. J. Burt and W. A. Lee, *A Family of Pyramid Structures for Multiresolution Image Processing*, David Sarnoff Research Center, Nov. 12, 1988.
11. P. J. Burt, Fast filter transforms for image processing, *Comput. Graphics Image Process.* **16**, 1981, 20–51.
12. S. Ranganath, Image filtering using multiresolution representations, *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 1991, 426–440.
13. D. Marr, *Vision*, Freeman, San Francisco, 1982.
14. H. R. Wilson, D. K. McFarlane and G. C. Phillips, Spatial frequency tuning of orientation selective units estimated by oblique masking, *Vision Res.* **9**, 1983, 873–882.
15. L. O'Gorman and A. C. Sanderson, A comparison of methods and computation for multi-resolution low and band-pass transforms for image processing, *Comput. Graphics Image Process.* **37**, 1987, 386–401.
16. D. Marr and E. Hildreth, Theory of edge detection, *Proc. Roy. Soc. London Ser. B.* **207**, 1980, 187–217.
17. A. M. Rohaly and G. Buchsbaum, Inference of global spatiochromatic mechanisms for luminance variations in contrast sensitivity functions, *J. Opt. Soc. Amer. A* **5**, 1988, 572–576.
18. A. M. Rohaly and G. Buchsbaum, Global spatiochromatic mechanism accounting for luminance variations in contrast sensitivity functions, *J. Opt. Soc. Amer. A* **6**, 1989, 213–317.
19. R. W. Rodieck, Quantitative analysis of cat retinal ganglion cell response to visual stimuli, *Vision Res.* **5**, 1965, 583–601.
20. C. Enroth-Cugell and A. W. Freeman, The receptive-field spatial structure of cat retinal Y cells, *J. Physiol.* **384**, 1987, 49–79.

21. R. A. Young, The Gaussian derivative model for spatial vision: I. Retinal mechanisms, *Spatial Vision* **2**(4), 1987, 273–293.
22. B. Levitan and G. Buchsbaum, Signal sampling and propagation through multiple cell layers in the retina: Modeling and Analysis with multirate filtering, *J. Opt. Soc. Amer A* **10,** 1993, to appear.
23. S. Chu and C. S. Burrus, Optimum FIR and IIR multistage multirate filter design, *Circuits Systems Signal Process.* **2**(3), 1983, 361–386.
24. M. Vetterli, Running FIR and IIR filtering using multirate filter banks, *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-36**(5), 1988, 730–738.
25. B. Levitan and G. Buchsbaum, Architecture-dependent properties of analysis multirate filter banks, *Proc. ICASSP*, 1991, 1805–1808.
26. B. Levitan and G. Buchsbaum, Conversions between parallel and hierarchic architecture analysis multirate filter banks, *IEEE Trans. Signal Process.* **40**(11), 1992, 2837–2841.
27. R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*, Prentice–Hall, Englewood Cliffs, NJ, 1983.

BENNETT LEVITAN received the B.S. (academic honors and distinction) in electrical engineering from Columbia University in 1985 and the M.S. degree in bioengineering from the University of Pennsylvania in 1987. He is presently pursuing a joint degree combining medicine and a Ph.D. in bioengineering at the University of Pennsylvania. He was awarded the Measey Foundation Combined Degree Scholarship in 1986 and the National Institutes of Health Medical Scientist Training Program in 1987. His research interests include multirate filtering theory and its application to retinal modeling, artificial life, and genetic algorithms. Mr. Levitan is a member of Tau Beta Pi, the Optical Society of America, ARVO, and IEEE.

GERSHON BUCHSBAUM received the B.Sc. and M.Sc. degrees in electrical engineering with distinction in 1974 and 1975, and the Ph.D. degree in 1978, all from the Tel Aviv University, Tel Aviv, Israel. Since 1979 he has been with the Department of Bioengineering at the School of Engineering and Applied Science, University of Pennsylvania, Philadelphia. In 1984 he was awarded the Presidential Young Investigator Award from the National Science Foundation. His research interests are concentrated in human vision, particularly image coding and analysis by the visual system. Dr. Buchsbaum is a member of IEEE, the Optical Society of America, and the Association for Research in Vision and Ophthalmology.

Group 3: Color constancy and interactions of space and color

Appendix G: Network Simulations of Retinal and Cortical Contributions to Color Constancy, Vision Research Vol. 35, pp. 413-434, 1995

Appendix H: A Multi-Stage neural network for Color Constancy and Color Induction, IEEE Transactions on Neural Networks, in press, 1995

Pergamon

0042-6989(94)00132-4

Vision Res. Vol. 35, No. 3, pp 413–434, 1995

0042-6989/95 $7.00 + 0.00

Appendix G

Network Simulations of Retinal and Cortical Contributions to Color Constancy

SUSAN M. COURTNEY,*† LEIF H. FINKEL,*‡ GERSHON BUCHSBAUM*

Received 10 May 1993; in revised form 28 December 1993; in final form 2 June 1994

A biologically-based neural network simulation is used to analyze the contributions to color perception of each of several processing steps in the visual system from the retina to cortical area V4. We consider the effects on color constancy and color induction of adaptation, spectral opponency, non-linearities including saturation and rectification, and spectrally-specific long-range inhibition. This last stage is a novel mechanism based on cells which have been described in V4. The model has been tested with simulations of several well known psychophysical color constancy and color induction experiments. We conclude from these simulations the following: (1) a simple push–pull spectrally specific contrast mechanism, using large surrounds analogous to those found in V4, is very effective in producing general color constancy and color induction behavior; (2) given some spatio-temporal averaging, receptor adaptation can also produce a degree of color constancy; (3) spectrally opponent processes have spatial frequency dependent responses to color and brightness contrast which affect the contribution of the V4 mechanism to color constancy in images with nonuniform backgrounds; and (4) the effect of the V4 mechanism depends on the difference between center and surround while the effect of adaptation depends on the total sum of inputs from both center and surround and therefore the two stages cooperate to increase the range of stimulus conditions under which color constancy can be achieved.

Color constancy Color induction V4 Adaptation

INTRODUCTION

Human color perception is not a simple function of the wavelengths of light reflected from a small area on a single surface. Instead, color depends on the spatial distribution of the wavelengths of light present in the entire image. The two most common phenomena which demonstrate this dependence are color constancy and color induction. Color constancy is the tendency of the colors of surfaces to remain more constant than would be suggested by the physical composition of the reflected light under changing illuminance conditions. It is thought that color constancy contributes to object recognition by allowing more reliable judgments about the object's surface properties regardless of the ambient light. A related phenomenon, color induction, is the change in the color of a surface due to its juxtaposition with other colored surfaces. Color induction enhances the color contrast in a scene and probably aids in object detection and surface segmentation.

Color constancy has been the subject of investigation for many years and a large variety of approaches have been attempted. Some were based on learning and judgment (e.g. Helmholtz, 1866; review by Jameson & Hurvich, 1989). Others have attempted to explicitly separate the reflectance from the illuminant by either computational theory (e.g. Buchsbaum, 1980; Maloney & Wandell, 1986; Rubin & Richards, 1982; D'Zmura & Lennie, 1986; Gershon & Jepson, 1989; Brainard & Wandell, 1991; D'Zmura & Iverson, 1993a, b) or linear filter theory (Faugeras, 1979). Additional well known theories include Land's Retinex (Land & McCann, 1971), various adaptation mechanisms (e.g. Hering, 1878; Helson, 1938; Judd, 1940; Brill & West, 1986; Brainard & Wandell, 1992), and spectrally-specific contrast based algorithms (Lucassen & Walraven, 1993). Most of these approaches attempted to identify one particular mechanism for achieving color constancy, or emphasized the importance of the contribution of one mechanism over another.

This emphasis has resulted in a retina vs cortex debate. Many researchers point to the need for two types of processing, one slow and one fast, one multiplicative and one subtractive (e.g. Hayhoe, Benimoff & Hood, 1987) to explain color constancy and color induction data. However, the different computational properties of these biological processes with regard to their effects on color constancy and color induction were not extensively studied, nor has much been said about the interactions

*Department of Bioengineering and the Institute for Neurological Sciences, University of Pennsylvania, 220 South 33rd Street, Philadelphia, PA 19104, U.S.A.

†Present address: National Institute of Mental Health, CPP, Building 10, Room 4C110, 10 Center Drive, MSC 1366, Bethesda MD 20892-1366, U.S.A.

‡To whom all correspondence should be addressed.

of these retinal and cortical processes. This paper attempts to examine these issues and to determine what advantage this multistage system has for producing color constancy.

Receptor adaptation and retinal spectrally opponent processes have been studied in depth, psychophysically, physiologically, and computationally, for their contributions to color processing and color coding. Interest in the cortical color mechanisms, particularly V4, has developed relatively recently and the results are more controversial. The first physiological evidence for the importance of the cortex in color constancy was reported by Zeki (1983) who recorded from individual cells in V4 whose responses, unlike those in V1, appeared to follow human color perception rather than wavelength. Several V4 lesion studies have had mixed results concerning the apparent effect of such lesions on color perception (Walsh, Kulikowski, Butler, & Carden, 1992; Heywood, Gadotti, & Cowey, 1992). Schein and Desimone (1990) reported that there are regions quite distant (up to 16 deg) from the classically-defined receptive fields of V4 cells which can influence a cell's response if the center of its classical receptive field is also stimulated. They called these regions silent surrounds. The existence of long-range lateral connections in V4 (Yoshioka, Levitt & Lund, 1992) and the dramatic reduction in ipsilateral surround suppression after section of the corpus callosum (Desimone, Moran, Schein & Mishkin, 1993) suggest that these large surrounds may be mediated by a mechanism within V4. The silent surrounds in V4 are sensitive to nearly the same wavelengths as the center of the receptive field, creating a spectrally-specific response which is functionally akin to "cone-specific contrast" (see Lucassen & Walraven, 1993). "Cone-specific contrast" appears, from psychophysical experiments, to be a necessary component of human color constancy (Tiplitz Blackwell & Buchsbaum, 1988b; Lucassen & Walraven, 1993; McCann, McKee, & Taylor, 1976). However, because of modifications to the cone inputs preceding cortical stages it is difficult to quantify the response of the V4 cells directly in terms of "cone-specific contrast".

Two psychophysical experiments, one using a split corpus callosum patient (Land, Hubel, Livingstone, Perry & Burns, 1983) and the other using binocularly fused stimuli (Shevell, Holliday & Whittle, 1992), demonstrate a significant influence from cortical processing in constancy and induction phenomena. In addition, regions significantly separated from the test area have been demonstrated by psychophysical experiments to be

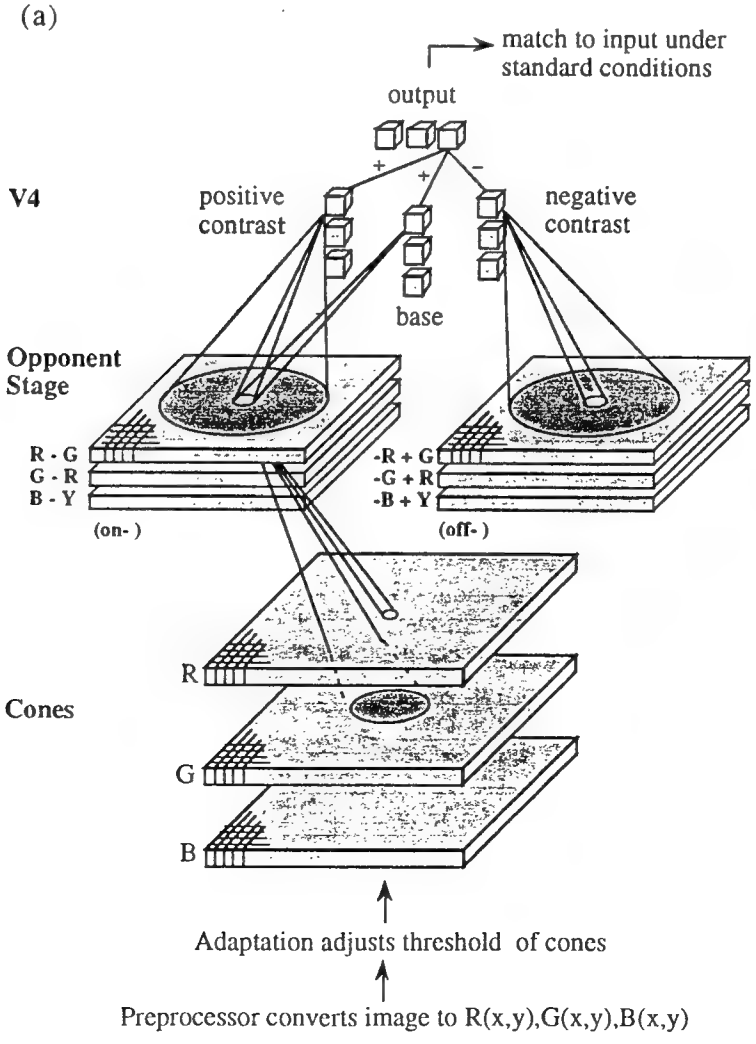


FIGURE 1.—*Caption on facing page.*

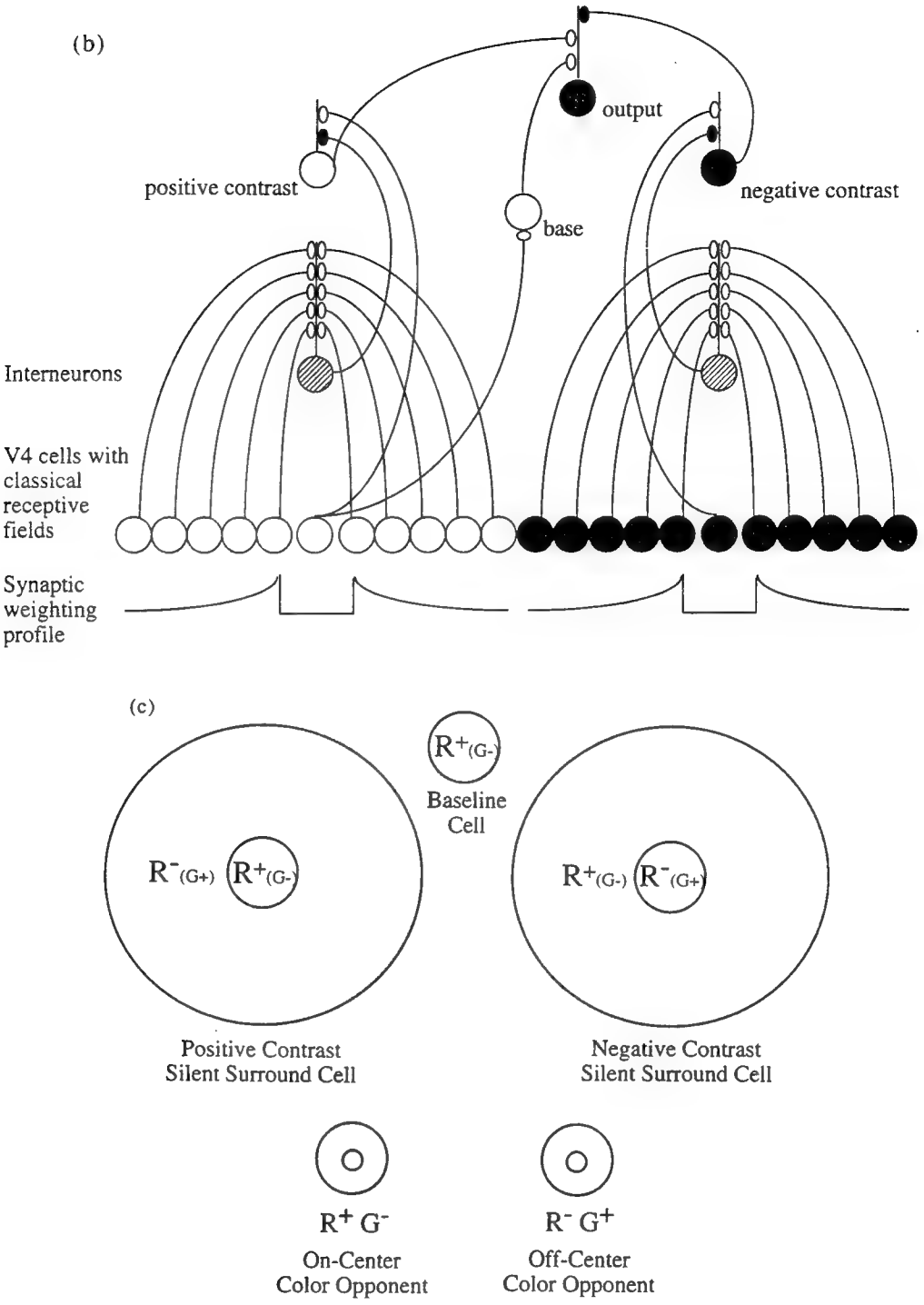


FIGURE 1. (a) Overview of entire model. Note the multiple, hierarchical stages of the network. Shaded regions show the connection fields of a single unit at each stage; lighter regions are excitatory connections, darker regions are inhibitory. The off-center spectrally opponent connection field, which is not shown, is the inverse of the on-center opponent connections. Units without silent surrounds in the spectrally-specific contrast stage receive only excitatory connections from on-center opponent units. (b) Proposed V4 push-pull mechanism. Detail of the cortical stages of the simulation. Open circles represent on-center cells, solid circles are off-center cells, and striped circles are interneurons. Synapses are shown in white for excitatory, black for inhibitory. The silent surrounds have an exponential synaptic weighting function as is shown at the bottom of the figure (c) Spatial structure of the receptive fields of the spectrally opponent and spectrally-specific stages. The figure shows only units with R centers, as an example.

very influential in determining perceived color (Tiplitz Blackwell & Buchsbaum, 1988a; Valberg & Lange-Malecki, 1990; Wesner & Shevell, 1992). The spatial dimensions of these phenomena are too large to be easily explained by known retinal structures. In addition, the speed with which a significant portion of this effect occurs, rules out the combination of receptor adaptation and eye movements as the sole mechanism for long-range color induction.

In order to explore the effects of both retinal and cortical processing on color constancy and color induction, we simulated a multi-stage neural network which includes three processes: receptor adaptation, spectral opponency, and spectrally-specific long-range inhibition. Each stage includes a saturating and rectifying nonlinear response function. Neural networks have been used before for implementing a variety of color constancy algorithms: lightness algorithms similar to Retinex (Hurlbert & Poggio, 1988; Moore, Allman & Goodman, 1991), a color categorization method using double opponent cells (Dufort & Lumsden, 1991), and an algorithm which uses contrast across boundaries to fill-in enclosed regions (Grossberg, 1987). In these simulations, as in other color constancy studies, the emphasis has been on describing a specific mechanism for achieving color constancy. In the current network simulation, which includes a new mechanism for cortical level processing, the specific effects of each processing stage and the interactions between processes were controlled and observed. We will show that a system which includes both retinal and cortical processes can produce the general behavior of both color constancy and color induction. In addition, we will demonstrate that while the differences between the spatial and chromatic properties of these processes sometimes leads to complex interactions between stages, all of these processes cooperate so that together they can produce greater contrast sensitivity and color constancy in a larger range of stimulus conditions than can any of the stages alone.

NETWORK ARCHITECTURE

An overview of the model is shown in Fig. 1(a). The cortical mechanism is shown in greater detail in Fig. 1(b). The network was simulated using NEXUS, an interactive neural simulator designed for large scale models (Sajda & Finkel, 1992). The complete network consists of over 11,000 cells and approx. 1.65 million connections. Below we will describe how each stage was implemented in the simulation. Table 1 summarizes the most significant parameters in the model.

(i) Input

The first stage corresponds to the cone responses. The input image is a 27×27 array, in which each entry defines the color at that location. The array is converted to three 27×27 arrays of cone activation levels: R, G, B. Therefore, an input image unit has a corresponding set of three units (analogous to one cone of each type) in the first layer of the network. Each entry in the input image is specified either by a Munsell reflectance spectrum and an illuminant spectrum, or in CIE notation (x, y, Y). When the reflectance and illuminant spectra were specified, the image was converted, at each point, to the three normalized cone activation levels by using the Vos–Walraven (Vos & Walraven, 1971; Vos, 1978) cone action spectra $[r(\lambda), g(\lambda), b(\lambda)]$, in steps of 10 nm:

$$R = \sum_{\lambda=400}^{700} k_1 r(\lambda)\mathscr{R}(\lambda)I(\lambda)\Delta\lambda$$

$$G = \sum_{\lambda=400}^{700} k_2 g(\lambda)\mathscr{R}(\lambda)I(\lambda)\Delta\lambda$$

$$B = \sum_{\lambda=400}^{700} k_3 b(\lambda)\mathscr{R}(\lambda)I(\lambda)\Delta\lambda \qquad (1)$$

where $\mathscr{R}(\lambda)$ is the reflectance spectrum, a fixed property of the surface, and $I(\lambda)$ is the illuminant, which may change with the particular viewing condition and, therefore, may change the (perceived) color of the surface. (Because inputs are computed from the reflectance and no other surface properties are considered, we will refer only to the reflectance spectra, not to a real or simulated surface.) The coefficients $k_{1,2,3}$ are constants which normalize the sensitivity spectra so that all cone types in the simulated array have the same peak sensitivity. Therefore, the three types of first layer units ("cones") have responses of the same order of magnitude and we designed the matching procedure to depend upon the relative responses of the three simulated color pathways [Section (vi)]. For those cases in which the image was specified in CIE notation, the image was converted to cone activation levels by applying the transformations for Vos–Walraven action spectra (Vos, 1978; Wyszecki & Stiles, 1982, p. 615) and then normalizing using the same coefficients $k_{1,2,3}$.

TABLE 1. Each of the most significant parameters in the simulation is presented along with the criteria used to determine that parameter's value (in parentheses are the specific values used and the range of possible values)

Parameter	Description	Factors in choice of parameter value
ω_{ij}	Connection strength between cells	Chosen to create receptive field shapes found physiologically, different for each cell type
σ_i	Threshold of cell i	Chosen so that most inputs fall in middle of response range, different for each cell type, cone threshold changes with adaptation state
β_i	Slope of linear portion of cell's response function	Chosen in combination with σ_i to give the appropriate dynamic range for each processing stage, different for each cell type
θ	Width of adaptation weighting function	Small value for fixation or very short presentation time experiments, large value for experiments with free eye movements ($\theta = 3.0$, relatively small compared to cortical silent surrounds, large compared to center of spectrally opponent receptive fields $0 < \theta <$ diameter of image)
α	Fraction of total long term adaptation achieved	Dependent upon length of viewing time ($\alpha = 0.2$; $0 \leq \alpha \leq 1$)
c_1, c_2	Coefficients for push pull mechanism	Chosen together with α to give a total average constancy shift of 20% in accordance with psychophysical data ($c_1 = c_2 = 0.25$; $0 \leq c_1 \leq 1$, $0 \leq c_2 \leq 1$)

(ii) Cell responses and nonlinearities

In the simulation of the network model, the total input to cell, Q_i, is determined by a weighted sum of the activities of all cells connected to cell i:

$$Q_i = \sum_{j=1}^{n} \omega_{ij} A_j \tag{2}$$

where A_j is the activity of cell j, ω_{ij} is the connection strength from cell j to cell i. The cells of the network corresponding to the cone layer have a Naka–Rushton response function (Naka & Rushton, 1966):

$$A_i = \frac{Q_i^x}{Q_i^x + \sigma_i^x} \tag{3}$$

where x is a constant from 0.7 to 1.0. In the simulation results shown here $x = 0.9$. The general behavior of the system was not very sensitive to the value of this parameter. σ_i is the threshold of cell i. The input, Q_i, for a cone is the cone activation level R, G, or B calculated from the input image as described above in equation (1).

In all other stages, cell activity is determined by a sigmoidal response function of the input:

$$A_i = (\max - \min)\left(\frac{1}{1 + \exp[-(Q_i - \sigma_i)\beta_i]}\right) + \min \tag{4}$$

where A_i is the activity of cell i, max and min are the maximum and minimum possible activity levels for cell i, σ_i is the threshold of the cell, and β_i is proportional to the slope of the linear portion of the curve (see Fig. 2).

(iii) Adaptation

We assume an initial long-term adaptation to a uniform neutral background [see Walraven, Enroth-Cugell, Hood, MacLeod, and Schnapf (1990) for a review of psychophysical and physiological studies on adaptation]. The amount of threshold shift $\Delta\sigma$, is determined by the difference between the cone activation level for the neutral background stimulus and the cone activation level for the new stimulus. Because adaptation is dependent on the temporally weighted average of its input, the adaptation shift for a cone is dependent not only on the point in the image directly corresponding to that cone position, but also on the surrounding area to which the cone may be exposed during eye movements, or from optical blur. We approximated this temporal effect by a two-dimensional Gaussian spatial weighting function, because for the psychophysical experiments we were interested in studying, there was generally either a fixation point, or a central test patch around which one could assume eye movements were centered. In the simulation, the amount of the shift follows a sigmoidal function of the difference between the neutral and the current stimuli and is proportional to the length of viewing time. These constraints are incorporated into the simulation by calculating the threshold shift for the receptor adaptation by using the equation:

$$\sigma_{new} - \sigma_{neut} = \alpha\left\{2M\left(\frac{1}{1 + \exp[-(Q_i - Q_{neut})\beta_i]}\right) - M\right\}$$

$$M = \left|\sum_{i=0}^{n} (Q_i - Q_{neut})\left(\frac{1}{2\pi\theta^2}\right)\exp\left[\frac{-(x^2 + y^2)}{2\theta^2}\right]\right| \tag{5}$$

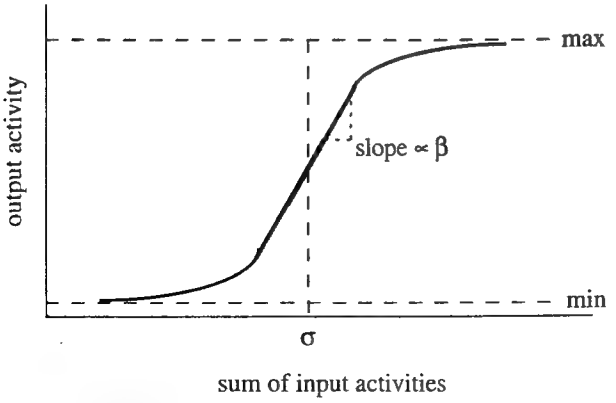


FIGURE 2. Nonlinear response function of each cell. The parameters are set for each stage so that most stimuli produce responses in the linear range of this function. The slope of the linear portion of the curve is proportional to β. Each cell's input is the weighted sum of the activities of all the cells connected to it. σ is the "threshold" which is defined for mathematical clarity to be at the center of the linear portion of the response. Saturation and rectification occur when the cell's output nears its maximum and minimum outputs respectively.

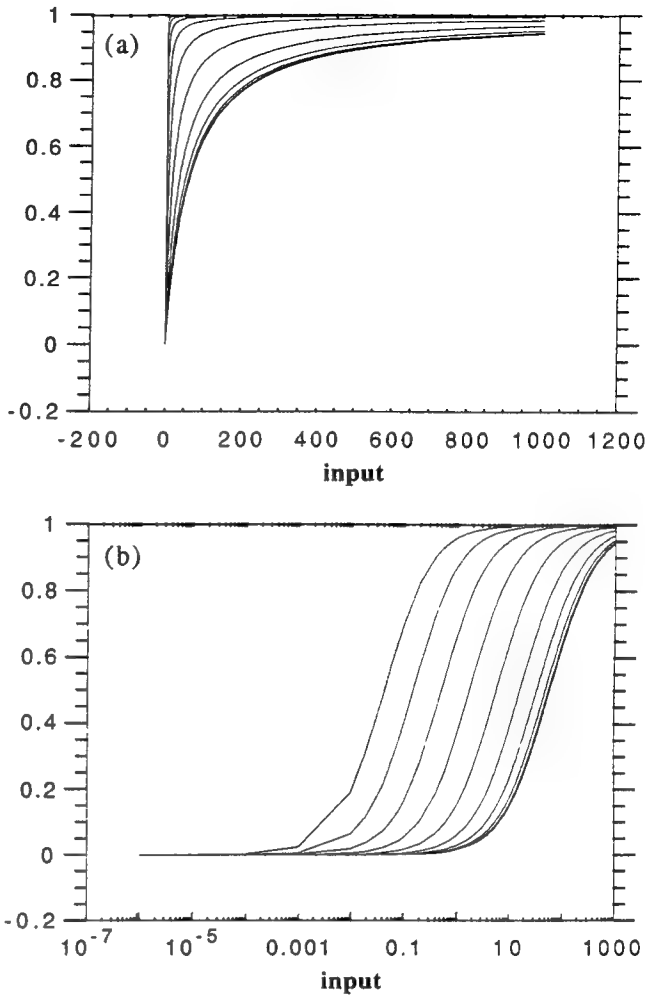


FIGURE 3. Response curves for cones in the simulation under a range of values for the adaptation threshold. (a) Shows the sigmoidal limits of the adaptation range. The luminance level of the adapting stimulus was increased linearly, but the threshold values reach an asymptote at both ends of the range. (b) Same as (a) in log–linear coordinates.

where σ is the threshold; β is the proportional to the slope of the linear portion of the function; Q_i is the cone activation level (i.e. R, G, or B) due to current image pixel i; Q_{neut} is the cone activation level due to standard neutral at image pixel i; n is the number of pixels in the image; x, y is the horizontal and vertical distances from pixel i to the center of the cone's receptive field when fixated on the center of the image; θ is the width of a Gaussian weighting function which varies with the degree of fixation required for the experiment; α is the fraction achieved within the stimulus presentation time of the total difference in long-term adaptation states between the neutral state and the state for the new stimulus.

α is proportional to the time of exposure. As α increases, the size of the threshold shift increases, following a sigmoidal curve ranging from $-M$ to $+M$ where M is the difference between the weighted average activation level for the current image and the activation level for a uniform neutral background (see Fig. 3). In the current study, α was held fixed at 0.3 and θ was held fixed at 3.0. However, we wished to include this flexibility in the model because eye movements do affect the adaptation state. With longer exposure time, the cell will be able to better adapt (larger α) to its new stimulus. Under certain experimental conditions, longer exposure time may also allow for more eye movements. The spatial extent of the weighting function broadens with more eye movements. In the extreme case of very long exposure time and completely random eye movements over the entire field of view, the weighting function would be flat and the cone would adapt to the field average. This dependence of the parameters θ and α on eye movements and viewing time, allows the effects of the adaptation stage of the simulation to vary with the experimental conditions being considered. This is important because the extent of eye movements in psychophysical experiments has been shown to affect color perception (Cornellissen & Brenner, 1991).

(iv) Spectral opponency

For the purpose of studying the effect of spectral opponency, we include only a single stage for this process, instead of the hierarchy of opponent cell types observed physiologically between the retina and V4. We wished to study the effects of spectral opponency as a mathematical operation rather than attempt to simulate the specific anatomical implementation. Opponency can occur at many levels of the visual system from cone gap junctions to the cortex (see review by Lennie & D'Zmura, 1988). We avoid the term "color opponency", because it has often been used in reference to psychophysical phenomena which may not necessarily be the result of spectrally opponent cells in a specific visual stage.

Opponent processing is achieved in the simulation by subtracting responses of spectrally opponent cone types and is generally based on the properties of LGN parvocellular type I receptive fields. In the simulation, each "cell" receives excitatory input from a single cone in the center of its receptive field and inhibitory input from several cone types surrounding the center using a difference of Gaussians synaptic weighting function (Lennie & D'Zmura, 1988). The surrounds receive input from all cones in their receptive fields, however the synaptic weights are different for each cone type. The surround input is most heavily weighted toward the cone type(s) opponent to the center cone type. For example, opponent cells whose centers receive excitatory input from R cones receive inhibitory surround input from both R and G cones, but the amplitude of the synaptic weighting function for the G cones is twice that for the R cones. The opposite ratio was used for the G center cells. The R and G centered cells, thus, do not differ from each other just by a negative sign, but have linearly independent cone input combinations. B center cells receive inhibitory input which is equally weighted between the R and G cones. Altogether there are three linearly independent combinations. Off-center cells were created by using the same weighting functions, but with opposite sign, and their thresholds were lower than those of the on-center cells, giving them a higher spontaneous activity level. Therefore, the off-center cells responses were greatest when the magnitude of the stimulus in the center of the receptive field was less than that in the surround. The off-center cells of course do not add additional independent combinations to the three resulting from the on-center cells. In addition, primate retinal and LGN cells do not have perfectly balanced centers and surrounds (Derrington & Lennie, 1984). Rather, the center strength (volume of two-dimensional Gaussian sensitivity profile) is roughly twice that of the surround, allowing these cells to have a significant response to homogeneous fields as well as to edges. Likewise, the spectrally opponent stage in the simulation has a 2:1 center/surround sensitivity ratio.

(v) Higher cortical processing

The next stage in the network is designed to respond according to the primary chromatic properties of the analogous cells in V4 (Schein & Desimone, 1990). These cells have large, suppressive surrounds each of which has a wavelength sensitivity similar to that of the center of the receptive field [see Fig. 1(b, c)]. These large surrounds had little or no effect on the cell's activity unless the center was also stimulated, and were therefore termed "silent surrounds". In the simulation, the "classical receptive field" (Schein & Desimone, 1990) receives excitatory input from a single class of spectrally opponent cells. These same type cells provide inhibitory input to the "silent surround" outside the classical receptive field. The "silent" behavior of the surrounds could be explained either by shunting inhibition (a multiplicative suppression of the excitatory input to a cell) or by rectified inhibition (the absence of effective inhibition in the resting state because of a very low spontaneous activity level). We chose to use rectified inhibition in the simulation because it is often found in cortical neurophysiological measurements while shunting inhibition appears to be rare in the cortex (Berman,

Douglas, Martin & Whitteridge, 1991). This is achieved by setting the thresholds of the silent surround cells so that the resting levels are very low. The effect of this rectification, together with the 2:1 center:surround weighting of the spectrally opponent cells, is to make the V4 cells in the simulation primarily dependent on the spectral sensitivity of the centers of the opponent cells which provide input to the cell. In this sense, the responses of the V4 cells in the simulation are measuring the difference in activity between the contributions of cones of the same type in the center and the surround. Therefore, we refer to the response of the V4 cells in the simulation as measuring spectrally-specific contrast.

Desimone, Schein and their colleagues (Moran, Desimone, Schein, & Mishkin, 1983; Desimone & Schein, 1987) reported that the effect of stimulation in the silent surround decreases with increasing distance from the classical receptive field. Psychophysical results also show a decrease in the effect of inducing regions with increasing distance (e.g. Tiplitz Blackwell & Buchsbaum, 1988a; Valberg & Lange-Malecki, 1990; Wesner & Shevell, 1992; Zaidi, Yoshimi, Flanigan & Canova, 1992). To incorporate these observations into the simulation, the inputs to the surround are weighted according to distance from the center by a negative exponential function [see Fig. 1(b)].

The strengths of the centers and silent surrounds of V4 cells appear to be well balanced; stimulation of the surround can completely inhibit the response to stimulation of the center (Schein & Desimone, 1990). Because the silent surround cells in V4 respond only when there is a difference, either in wavelength or luminance, between the center and the distant surround, these cells are particularly well suited for carrying information about contrast. However, for those images that have little spectrally-specific contrast, or an unknown or non-gray average chromaticity (e.g. blue sky, green forest), the d.c. (or local average chromaticity) information is also important. It is significant, therefore, that approx. 10% of the cells found in V4 did not have silent surrounds. The cells without silent surrounds have the same classical receptive field response as those cells with silent surrounds. These cells have the capacity to carry the (spatial) d.c. portion of the signal, i.e. to respond to homogeneous fields as well as edges and small spots. These center-only cells have been included in the network and we refer to them as "local reference cells" because they provide the normalizing reference information for the contrast cell responses.

The responses of analogous V4 stage "cells" in the simulation were created directly using the outputs of the spectrally opponent stage. A positive contrast cell receives its input, excitatory from the center and inhibitory from the surround, from on-center spectrally opponent cells. Therefore, the positive contrast cells respond to images for which the input to its classical receptive field is greater than the input to its silent surround. We have also included negative contrast cells which receive input from off-center cells, and therefore respond when the center input is less than the surround input. While, to our knowledge, there has been no systematic study of off-center cells in V4, given the symmetry of on- and off-populations of cells in earlier stages and the common observation that color constancy and color induction are seen in negative as well as in positive contrast stimuli, it seems reasonable to propose a negative contrast cell analogous to the positive contrast silent surround cells. Alternatively, the functions of both the negative and positive contrast cells in the simulation could be achieved by the V4 cells, also described by Schein and Desimone (1990), which had silent surrounds with both spectrally-specific inhibition and spectrally opponent excitation.

In order to combine the physiological information from the local reference and contrast cells into a simple set of outputs which could be compared to human color perception, we combined the outputs of these V4-like cells into a simple push–pull mechanism. [This stage is shown in Fig. 1(b).] We used one reference cell for every pair of positive and negative contrast cells. The output of this final network stage is determined by the response of the local reference cells, enhanced by the positive contrast cells, or inhibited by the negative contrast cells. This is given by the equation:

$$O = B + c_1 P - c_2 N \tag{6}$$

where O is the output, B is the local reference response, P is the positive contrast response, N is the negative contrast response, and c_1 and c_2 are constants. The constants c_1 and c_2 were chosen, together with α to give an average constancy shift of 20% of the distance between the color of the reflectance under the standard illuminant and the color of the reflectance under the test illuminant. This is consistent with psychophysical data (Tiplitz Blackwell & Buchsbaum, 1988b).

(vi) Matching procedure

After the image was processed by these three model stages, we needed to assess the input–output relationship in a manner similar to the psychophysical experiments. We, therefore, used a process analogous to the psychophysical matching paradigm (see Fig. 4). The final

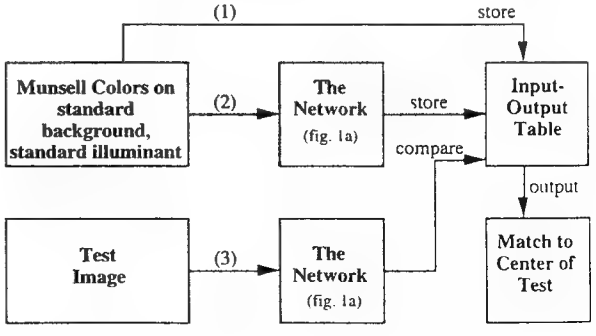


FIGURE 4. Block diagram of the matching procedure. Paths (1) and (2) are the storage procedure and take place simultaneously. Repeating this process results in the storage in memory of a lookup table of input/output pairs for 2625 evenly spaced colors under standard background and illuminant conditions. Once this is complete, the output of the network for a test image is compared to the outputs for the standard images and the closest match is determined (path 3).

output is a single set of three cells whose receptive fields are centered on the middle of the input image. (Because the sizes of the receptive fields increase with each subsequent stage in the network model, the dimensions of the network layers decrease progressively in order to reduce edge effects.) If the outputs for two different images are equal, then the centers of the two images are said to "match".

In order to do this matching efficiently, for each set of simulation parameters, outputs were determined for 2625 colors [from Table I(6.6.1) of Wyszecki and Stiles (1982) which lists CIE coordinates for Munsell colors] using a standard background and illuminant. Unless otherwise indicated, reported matches were made using calculated input images corresponding to a single small square (3×3 input units) of a Munsell reflectance against a uniform gray (Munsell N6.0) background under CIE standard illuminant *C*. These standard outputs are then stored with their corresponding input values in a look-up table. Then, when the test image is shown to the network, its output is compared to the stored outputs for the standard images. The standard input color which corresponds to the stored output closest to the test image output is reported as the "match".

SIMULATION RESULTS

(i) General constancy and induction abilities

We tested the network with various stimuli to determine how well it would follow human perception in the primary aspects of color constancy and color induction. The first simulated experiment tested brightness constancy and brightness induction. The center of the image was a single small patch (3×3 units) of the gray Munsell reflectance N6.75. The background of the first test image was Munsell reflectance N6.0. Constancy was tested using several different luminances of a spectrally flat illuminant. Matches were made using a N6.0 background and CIE standard illuminant C which gives a luminance of approx. 43 cd/m^2 for the N6.75 reflectance. Therefore, for the N6.75 center reflectance under other illuminants, perfect luminance constancy would be achieved if the matches also had a luminance of 43 cd/m^2.

The results are shown in Fig. 5. Because the chromatic changes under these conditions were small, only the luminance results are shown. The input luminances are shown by the black columns. The gray columns represent matches to the center of the first test image under the different levels of illuminant. For the N6.0 surround condition (the same surround used for the match condition), the match luminances were equal to the physical luminance (43 cd/m^2) of N6.75 under the standard illuminant for all the different test illuminants, demonstrating brightness constancy. For the second test image a lighter background, N7.5, was used. The matches for this image are shown by white columns. Again, all illuminant conditions produce matches of the same luminance, demonstrating brightness constancy. However, the presence of the lighter surround shifts the luminance matches to a smaller value, the correct shift direction for brightness induction.

For color constancy, 10 different colored reflectance patches were used with three different illuminants. The reflectances were chosen, one of each Munsell hue, as a representative sample of Munsell chips of moderate luminances. One illuminant peaked at 440 nm, one at 560 nm and one at 660 nm. Again CIE standard illuminant *C* was used for the match condition. For both the match and test images, the background was Munsell reflectance N6.0. For most of the reflectance–illuminant pairs (25 out of 30), some degree of color constancy was obtained by the network. Figure 6 shows results for two of the 10 reflectance patches under the three colored

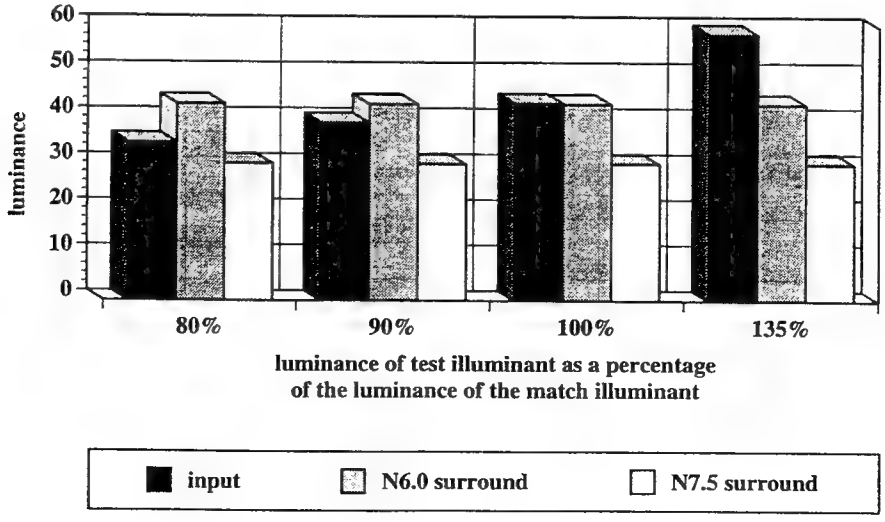


FIGURE 5. Demonstration of the network's ability to do both brightness constancy and brightness induction. The luminances of the illuminants used are given as a percentage of the luminance used for the matching condition. When the surround reflectance is the same as in the match condition (N6.0) the luminances of the matches are all equal to the physical luminance of the test patch under the standard illuminant, demonstrating brightness constancy. When the surround reflectance is a higher luminance, all of the matches shift to a lower luminance, demonstrating brightness induction.

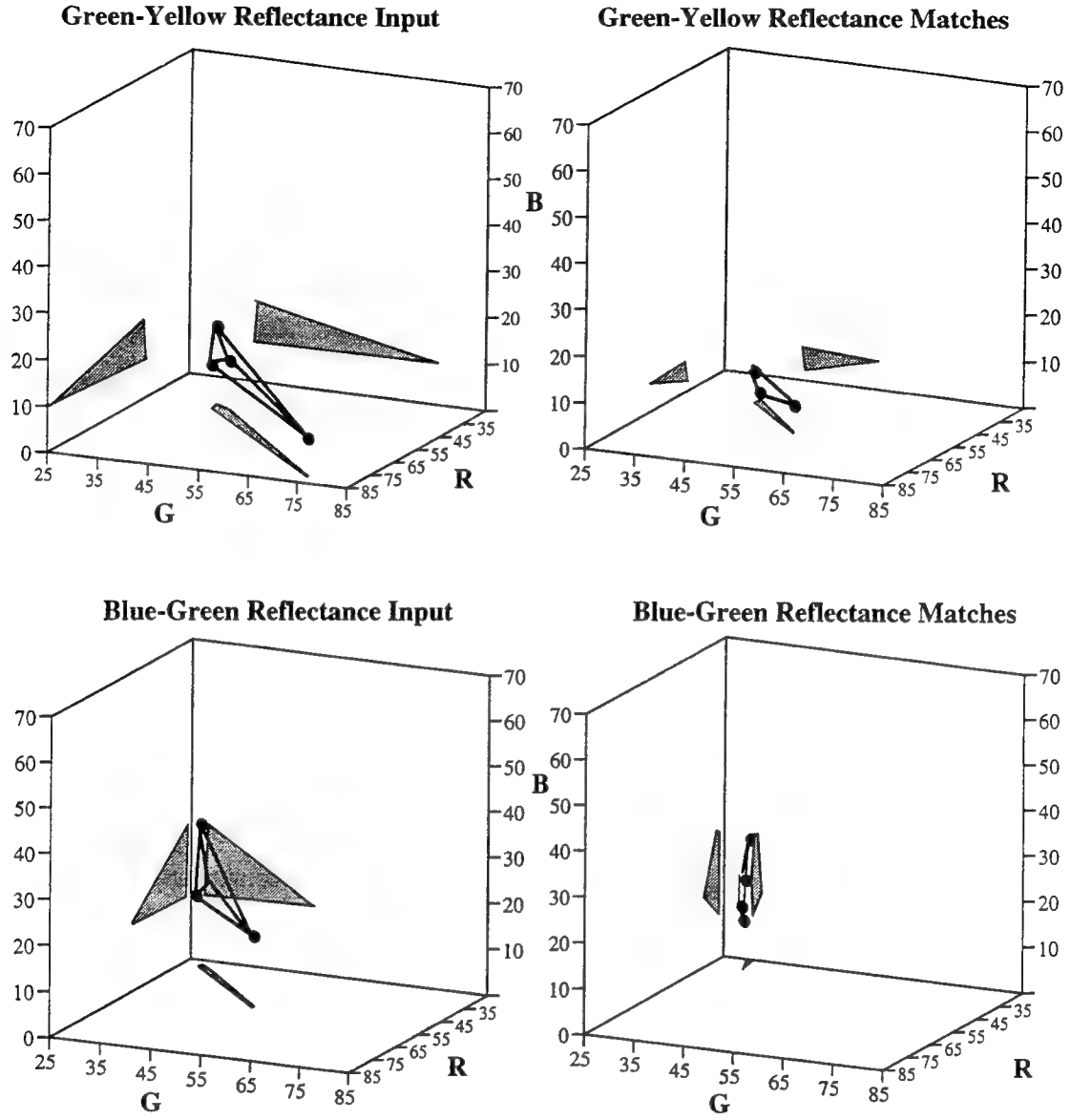


FIGURE 6. Two examples of the network's color constancy ability. The left two graphs show input values (cone activation levels: R, G, B) for two reflectances under four different illuminants (peaking at 440, 560, and 660 nm, and illuminant *C*). The right two graphs show the matches to those inputs. If the network showed perfect color constancy, all the matches for a single reflectance would be at a single point. If the network showed no color constancy, the matches would be identical to the inputs. (Note that in some graphs, plotted points superimpose.)

illuminants and illuminant *C*. A match is considered as "*achieving some degree of color constancy*" if the difference (in color space) between the color of the match and the "true color" is less than the difference between the "true color" and the "physical color". Both "true color" and "physical color" are defined by their computed coordinates in the RGB space described earlier. "True color" is the computed coordinates of the reflectance under standard illuminant conditions, and "physical color" is computed coordinates of the reflectance under the test illuminant. A "*shift toward constancy*" is a shift of the match toward the true color and away from the physical color. In the color constancy tests, the matches made by the network are somewhat color constant, but do not completely compensate for the illuminant change. The size of the constancy shift is different for each reflectance–illuminant pair and the amount of compensation can be varied by changing α, c_1, and c_2. However, we were not able to achieve perfect color constancy for all stimuli with any of the parameter combinations that we tried. This is not unexpected since human color "constancy" is also imperfect (see review in Beck, 1972; Tiplitz Blackwell & Buchsbaum, 1988b).

In a second test of color constancy, we simulated the McCann Mondrian experiment (McCann, McKee & Taylor, 1976). The experimental set-up is shown in Fig. 7(a). Two identical Mondrians were simulated, one under a standard neutral illuminant (CIE illuminant *C*) and the other under a combination of illuminants chosen so that the center colored patch [purple-blue in the example shown in Fig. 7(b)] under that illuminant would have the same *R*, *G*, *B* as a gray (N7.5) patch under the standard illuminant. Matches were made using the Mondrian as the background rather than the neutral

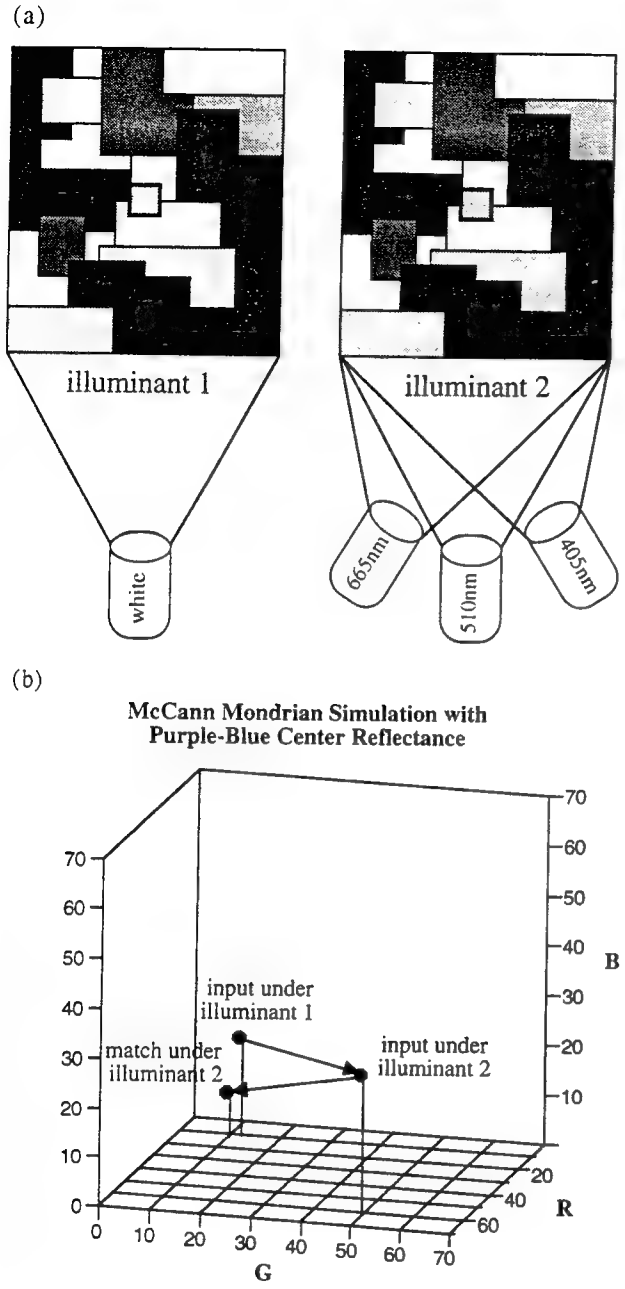


FIGURE 7. (a) McCann Mondrian set-up for the psychophysical experiment (McCann *et al.*, 1976) and for the present simulation. The two Mondrians are identical except for the reflectance of the center patch. The first one is under a neutral illuminant (illuminant 1). The second is under a combination of three illuminants whose luminances have been adjusted so that the R, G, B values (cone activation levels) of the center patch in that Mondrian are equal to the R, G, B values for a gray reflectance under the neutral illuminant (illuminant 2). The central patch in the first Mondrian (under the neutral illuminant) is then chosen to "match" the (perceived) color of the center patch in the Mondrian under illuminant 2. (b) The model's results for the simulated Mondrian constancy experiment. The network demonstrates a shift toward constancy in both color and brightness by moving away from the "physical color" (input under illuminant 2) and towards the "true color" (input under white reference illuminant 1).

uniform field used in the other simulations [see Fig. 7(a)]. The color chosen for the center patch of the Mondrian under the standard illuminant to match the center of the Mondrian under the second illuminant, again, showed a tendency toward constancy, but not perfect compensation. For perfect constancy, the match would have to be identical to the color of the test patch under neutral illumination. For no color constancy, the match would have been equal to the color of the gray patch under neutral illumination.

Next, to test the spatial properties of color induction, we used small (3×3 units) reflectance patches surrounded by an annulus the width of which varied from 0 to 4 input units. The center patches and the surrounding annulus were separated by a neutral gap of 0 to 4 units in width. The diameter of the V4 surrounds in the simulation was 11×11 input units. Beyond the annulus, the background was the same neutral as the gap. The stimulus is shown in Fig. 8(a). As the width of the gap was increased, the amount of induction decreased [see Fig. 8(b)]. When the gap was 4 units wide, the annulus was outside the receptive field of the V4 cells and there was almost no induction. The induction effect did not disappear in the presence of a small gap as it would with a contrast mechanism which was highly localized. In addition, if the gap width is fixed and the width of the annulus is increased, the amount of induction increases [see Fig. 8(c)]. These results agree with those presented for the analogous psychophysical experiment by Tiplitz Blackwell and Buchsbaum (1988a).

The observation that induction is still noticeable when a neutral gap separates center and annulus, suggests that this same, large, spatially distributed spectrally-specific contrast mechanism could also account for the color context effects in psychophysical experiments by Wesner and Shevell (1992) in which they demonstrated that local contrast alone could not entirely account for color appearance. Wesner and Shevell used monochromatic lights to test color context effects, using the color cancellation method for a unique yellow center. The results for a simulation of these experimental conditions are shown in Fig. 9. The stimulus is shown in the figure inset. The stimulus used for the simulation consists of a central test spot (3×3 input units), an adjacent surrounding annulus (1 unit wide), and a distant surrounding annulus (3 units wide) immediately outside the adjacent annulus. The simulation was done using matches instead of cancellation, but the general results are the same. The results show that both areas adjacent to the test spot and distant areas affect the predicted color match. Green in either the adjacent or distant surround shifts the appearance of the yellow center toward red. Red in the distant surround shifts the appearance of the center toward green. Increased luminance of the test spot relative to the surround luminance decreases the induction effect.

(ii) The roles of V4 and adaptation

To understand what each stage contributes to color constancy and color induction, we repeated several of these simulated experiments with various stages in the network eliminated or modified. By eliminating the adaptation stage, we found that many of the general properties of color constancy could be achieved by the cortical spectrally-specific push-pull mechanism alone.

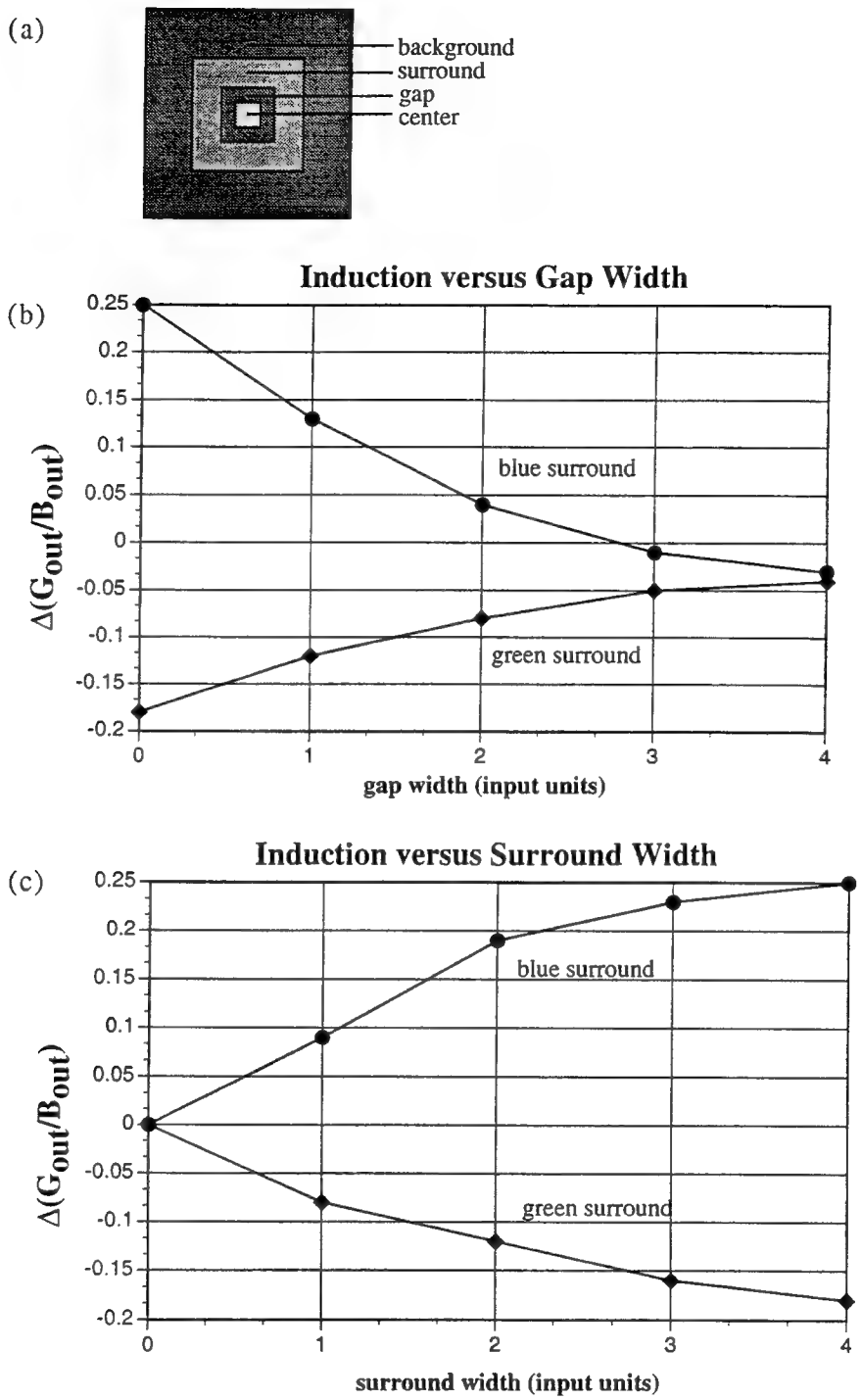


FIGURE 8. (a) Simulation image used to measure the spatial properties of color induction. Width of the gap and of the surround can be varied. The center patch is 3 × 3 input units, the same size as the centers of the V4 receptive fields. (b) Results for one example of color induction. The center reflectance is blue-green. A blue surround increases the $G:B$ ratio for the matches (circles), while a green surround decreases the ratio (diamonds). As the gap between center and surround is increased, the induction effect decreases. At 4 units separation, the surround of the stimulus is outside of the silent surrounds of the V4 cells. (c) With no gap, the width of the surround is varied. As the width of the surround increases, the amount of induction increases

Likewise, for most reflectance–illuminant pairs, adaptation alone also results in some degree of constancy. An example is shown in Fig. 10. However, each of these stages works best in different situations.

The size and direction of the color shift depends on stimulus conditions. An illuminant which causes a larger shift in the color signal (reflectance times illuminant) for the center test patch than for the surrounding background, such as a red test spot on a neutral background under red illuminant, causes an adaptation shift in the direction of constancy. However, the V4 stage for this stimulus increases the illuminant's influence rather than decreasing it because the R cone contrast is positive. On the other hand, an illuminant which causes a larger shift in the background than in the test spot, such as a blue illuminant on a yellow test spot with a neutral

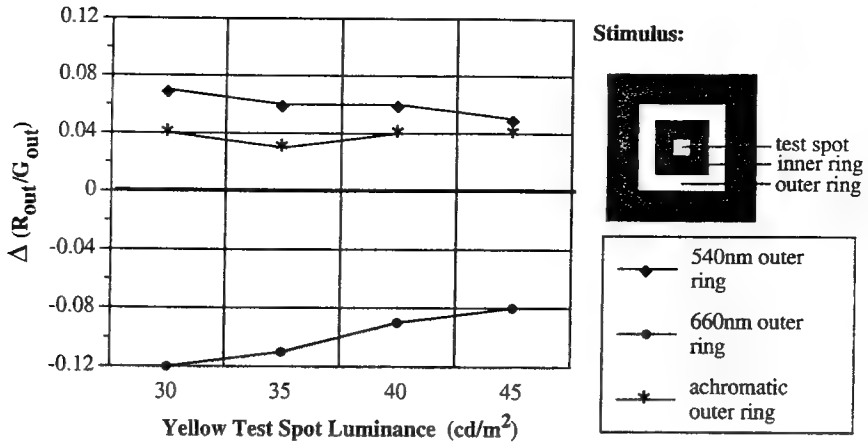
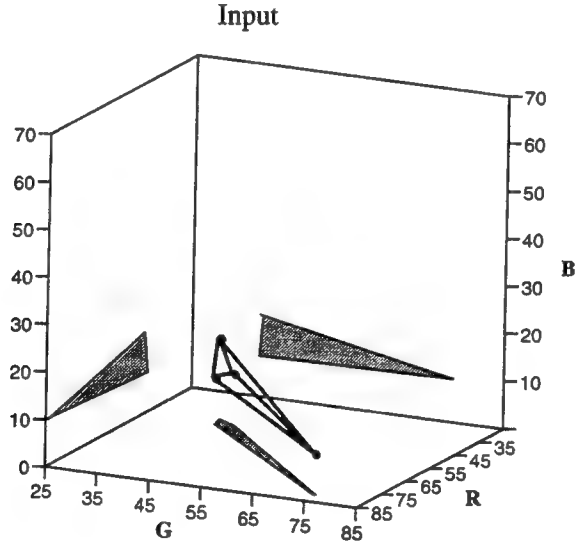


FIGURE 9. Results for the simulation of a color context stimulus, shown at the top right. The test spot is yellow, the adjacent surround is green (540 nm), and the distant surround is either green, red, or white. The figure shows the change in the (R/G) ratio from the neutral surround condition to the match for the yellow spot with the various colored surrounds. The results show that both areas adjacent to the test spot and distant areas affect the predicted color match. Green in either the adjacent or the distant surround increases the R/G ratio of the match, while red decreases the ratio. Increased luminance of the test spot relative to the surround luminance decreases the induction effect.



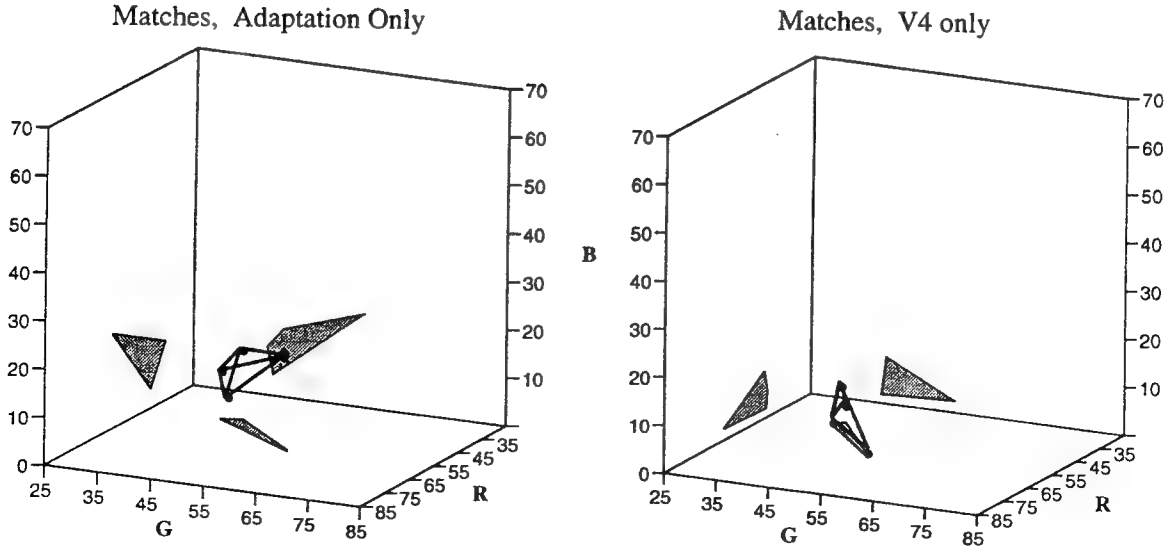


FIGURE 10. Color constancy simulation results plotted as in Fig. 5. The left graph shows the matches obtained when the V4, spectrally-specific contrast stage is eliminated and, therefore, only adaptation contributes to color constancy. The right graph shows the matches when adaptation is eliminated and, therefore, only the cortical spectrally-specific contrast mechanism contributes.

background, results in a large shift toward constancy at the V4 stage. In fact, the V4 stage overcompensates, causing color induction. The adaptation stage contribution depends on the degree of localization. Highly localized adaptation (perfect fixation) in this case results in almost no constancy shift, while less localized adaptation does cause a shift toward constancy.

The main reason for the difference in the color constancy contributions of these two stages can be seen in the spatial sensitivity profiles of each mechanism (see Fig. 11). The adaptation stage sums its input across both the test spot and the background; it is not spatially opponent. Whether most of the contribution is from the test spot or the background depends on how localized the adaptation is. The cortical contrast cells, on the other hand, receive antagonistic inputs from center and surround. Therefore, the effect of the V4 mechanism will depend on the difference between center and surround while the effect of adaptation will depend on the sum of inputs from both center and surround.

In situations where both stages, separately, would be effective in producing color constancy, their effects are sometimes antagonistic. Localized adaptation can decrease the contrast of the inputs to the center and surround of the V4 cells, making the V4 stage less effective. In some such cases, the size of the constancy shift with both stages is actually less than for either stage alone. However, the multi-stage system is more

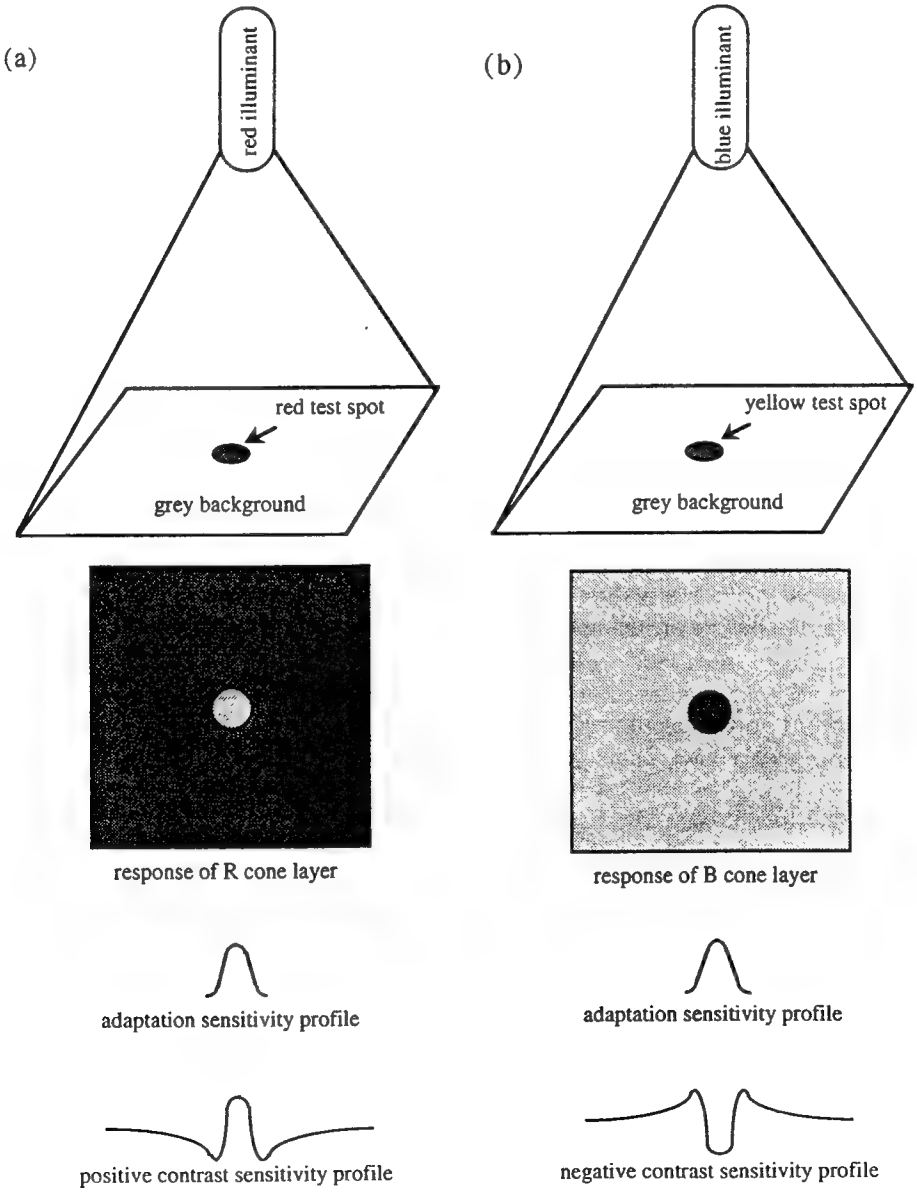


FIGURE 11. Two stimulus conditions, each of which favors a different mechanism in the network for achieving color constancy. (a) A red spot will reflect a red illuminant more strongly than will a gray background of equal lightness. Therefore, the adaptation mechanism, which is most sensitive to the test spot will respond well to the red illuminant and provide good color constancy. The spectrally specific contrast mechanism, on the other hand, has a positive contrast response, and therefore, enhances rather than diminishes the effect of the illuminant. (b) The opposite stimulus condition. The blue illuminant is reflected best by the background. This leads to little response from the adaptation mechanism but a good response from the negative contrast cells in the final layer of the network.

consistent than either stage alone, because for the cases in which one of the stages alone would fail to produce constancy, the other stage can generally compensate. We tested the network with 10 colored test spots on a neutral background under 3 illuminants, as described earlier. Without the V4 stage, the system shifts the match toward constancy for 20 of 30 stimuli. Without the adaptation stage (but with V4) the system succeeds for 22 out of 30. With all stages included, the system achieves some degree of constancy for 25 out of 30 stimuli. For four out of the five stimuli for which the complete system does not achieve constancy, neither adaptation alone nor the V4 mechanism alone could produce constancy. The complete system is capable of producing color constancy in a broader range of stimulus conditions than can be handled by either stage alone. The complete system, therefore, also has a slightly better *average* color constancy performance.

We wanted to have some quantitative measure with which to compare the amount of constancy achieved by each of the stages in the model. Although $(R^2 + G^2 + B^2)^{1/2}$ (where R, G, and B are the normalized cone activation levels) cannot be considered a true measure of "color distance" because R, G, and B are not orthogonal and also because the "distances" do not correspond to perceptual distances, it is a good intuitive measurement and incorporates both color and brightness. CIELUV color differences, ΔE^*, is a less intuitive measure, but one which does correspond to perceptual distances (Wyszecki & Stiles, 1982, p. 166). ΔE^* was also computed for each input-match pair and these numbers gave similar results. Figure 12 shows histograms of the $(R^2 + G^2 + B^2)^{1/2}$ "distances" from the actual matches under various colored illuminants to the ideal constancy match. The combination of both adaptation and V4 results in both a slightly smaller mean distance and a smaller range of distances than either stage alone.

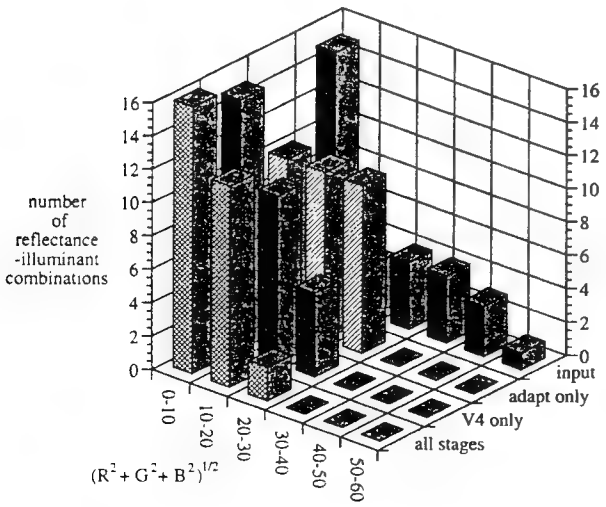


FIGURE 12. Histograms of the distances of each of the matches from perfect constancy. Distances are plotted separately for input, and for the network matches with adaptation only, V4 stages only, and all stages active. The plots show that while adaptation alone and V4 alone are each effective in reducing the largest color differences, all of the stages working together are able to achieve a lower average distance, and therefore "better" color constancy, than either stage alone.

(iii) Spectrally opponent vs spectrally-specific stages

The spectral sensitivities and center-surround organization of receptive fields in the opponent stage modify the inputs to the spectrally specific cortical stage. The effect that this intermediate stage has on the final output depends on the spatial structure and spectral composition of the input image (i.e. the segment sizes, spatial frequency content, number of edges, amount of chromatic and luminance contrast at the edges). In the following section we examine the effects that the opponent stage has on the input signal that it provides to the final stage of the network, and the effect that these modifications have on the output of the network.

Responses to high and low spatial frequency stimuli. If a low spatial frequency input (such that center and surround of the receptive field receive approximately the same input) to a spectrally opponent R–G cell changes in color, from yellow to red, without changing in luminance, the cell will receive both an increase in excitation and a decrease in inhibition. Spectral opponency, therefore, results in a high gain for low spatial frequency purely chromatic signals. On the other hand, the response to a low spatial frequency luminance stimulus will be attenuated because the increase (or decrease) in excitation will be offset by the increase (or decrease) in inhibition. At high spatial frequencies, this response relationship is reversed for cells which are spatially as well as spectrally opponent. A cell whose inhibitory surround falls partially on the darker side of a luminance edge will receive less inhibition than a cell which has both center and surround receiving input entirely from the higher luminance region (see Fig. 13). This center–surround receptive field structure, therefore, leads to enhancement of the cells responses to luminance edges. This is shown by the response of the spatially opponent cells in the network simulation.

On the other hand, at an equiluminant chromatic edge, a spectrally and spatially opponent cell may receive more inhibition from a surround which receives input partly from the other side of the color edge, if the surround is more sensitive to that color. As the response of the spectrally and spatially opponent layer of the network shows, this increase in inhibition results in a blurring of the chromatic edge response, an attenuation of high chromatic spatial frequencies (see Fig. 13). Because the V4-type spectrally-specific contrast cell has a very large receptive field, both the high and low spatial frequency responses of the spectrally opponent cells, which comprise the input to the V4 cell, are linearly summed.

The effect on color induction There has recently been some discussion in the field regarding the effect of the image spatial structure on color induction. Valberg and Lange-Malecki (1990) presented evidence that the color induction shift caused by a Mondrian background was the same as the induction caused by a homogeneous background whose chromaticity and luminance were equal to the spatially weighted average of the Mondrian background. This homogeneous background was termed

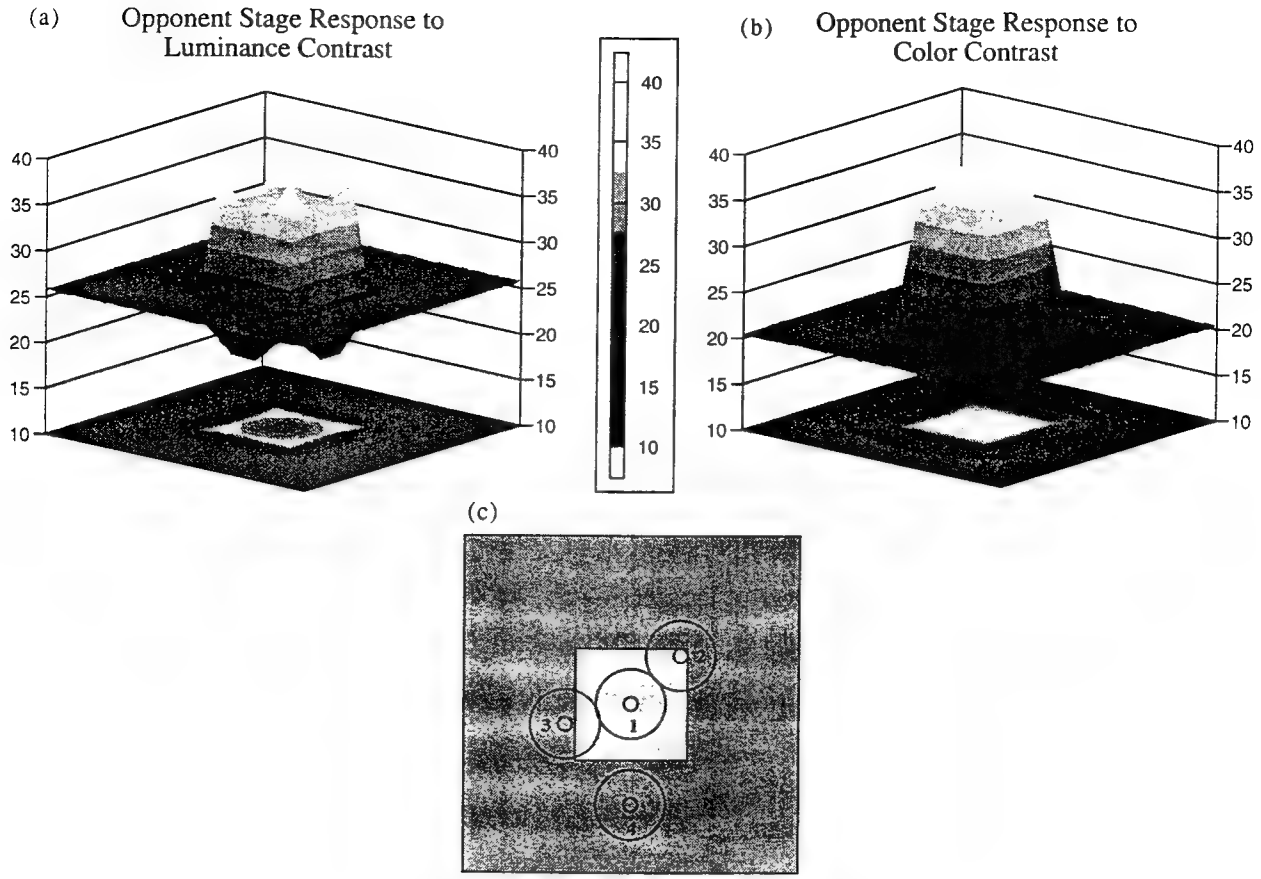


FIGURE 13. (a) Response of the R on-center spectrally and spatially opponent network layer to a light gray square on a dark gray background. The response to luminance contrast shows edge enhancement, unlike the response to an equiluminant color edge, shown in (b). Response of the R on-center spectrally and spatially opponent network layer to a red square on an equiluminant yellow background. The response to color contrast shows blurring at the edges, demonstrating low pass filter behavior. (c) The location of a cell's receptive field relative to a color or luminance edge affects its level of response. In the luminance contrast stimulus a cell with its receptive field at location 2 would have a greater response than a cell at location 1 because cell 2 would receive less input to the inhibitory portion of its receptive field. Similarly, a cell at location 3 would have a smaller response than a cell at location 4 because cell 3 would receive more inhibition. In other words, for luminance contrast resp2 > resp1 > resp4 > resp3, while for color contrast resp1 > resp2 > resp3 > resp4.

the "equivalent surround". Two additional psychophysical studies have since shown that perhaps the equivalent surround calculation must include some edge enhancement before the spatial average is computed (Brown, 1993; Wesner & Shevell, 1993). Because the opponent stage of the network causes luminance edge enhancement and the silent-surround stage calculates a spatially weighted average, we expected the simulation to show similar behavior.

We tested this hypothesis with the current simulation by using several input images whose surrounds had identical average color and luminance properties, but had an increasing number of high frequency edges. The stimuli and the results are shown in Fig. 14. The equivalent surround hypothesis predicts that such surrounds would have identical induction effects on the center test patch. The simulation outputs showed only very small changes with increased number of edges in the surround if the edges were purely chromatic. As explained above, a chromatic edge is not enhanced by the spectrally opponent cells. There was also no significant change when the luminances and saturations of all regions in the surround were such that all of the opponent cells were operating in the linear range of their response functions. However, there was a change in the output when some of the regions in the image produced responses outside the linear range of the opponent cells. For these images, the edge enhancement caused by the opponent cells was not symmetric across the edges. Therefore, the spatially weighted surround calculated by the silent-surround cells was different for each of the different images.

The effect on color induction of high spatial frequencies in both color and luminance has also been shown psychophysically in a different paradigm. Zaidi *et al.* (1992) showed an attenuation in the magnitude of color induction when an equiluminant surround included high spatial frequencies. Shevell and Wesner (1990) found a larger decrease in the magnitude of color induction when a thin white ring, equiluminant with the surround color, was placed in the surround, than when a black ring was placed in the surround. Zaidi *et al.* (1992) argue that this could also be explained by an attenuation of color induction by high spatial frequency chromatic signals in the inducing surround. However, neither Zaidi *et al.* (1992) nor Wesner and Shevell (1990) found this

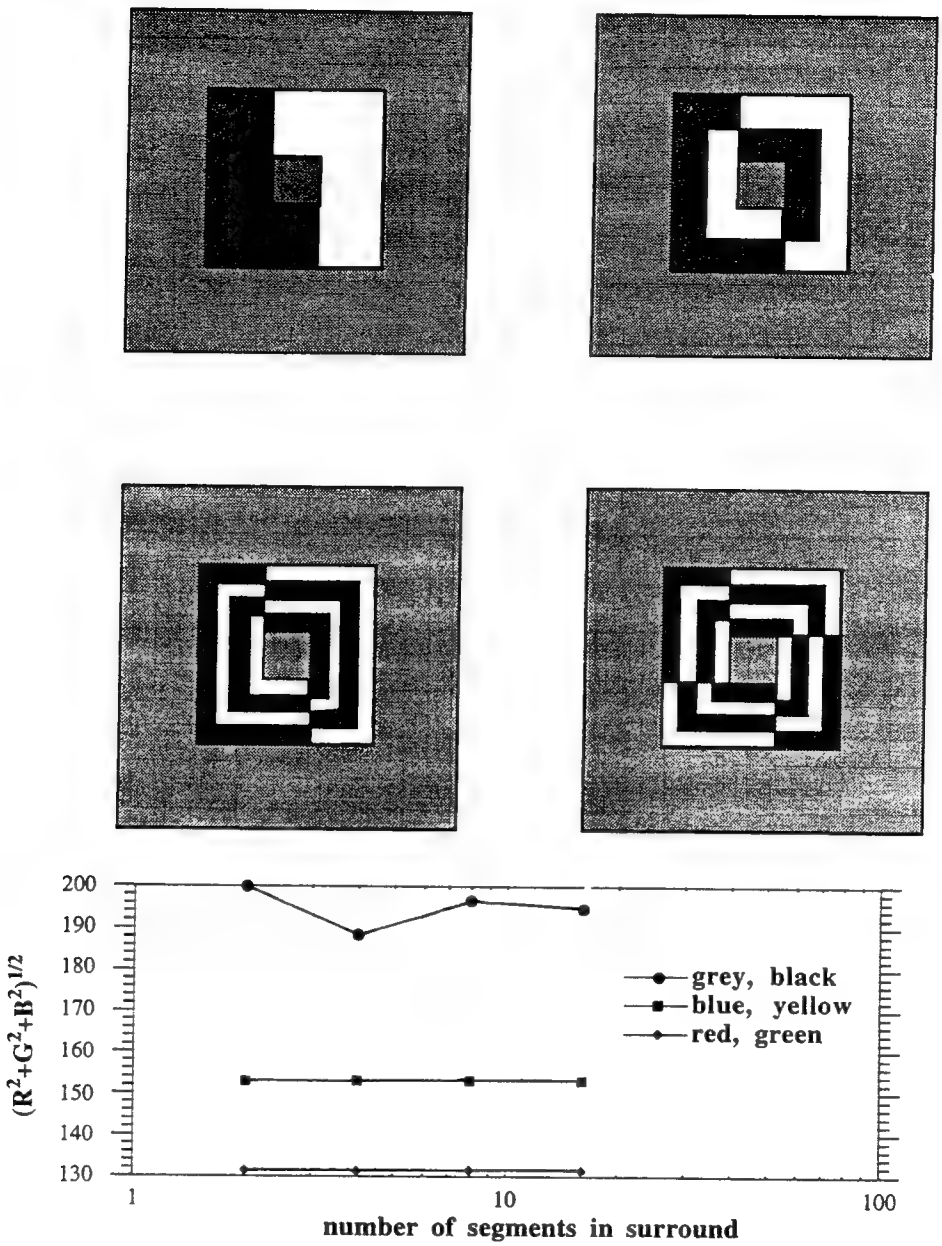


FIGURE 14. Four stimuli are shown, the surrounds of which all have identical spatial averages. The network responses for these stimuli, however, are not always identical. The sum of the squares of the outputs is shown for each of the four stimulus types with various colors assigned to the sectors in the surrounds. The outputs for blue–yellow surrounds [shown by squares, inputs (R, G, B) = (24.2, 17.9, 9.8) and (24.2, 17.9, 79.8)] and for red–green surrounds [shown by diamonds, inputs = (44.2, 32.9, 19.8) and (24.2, 22.9, 19.8)] showed only very small changes when the spatial structure of the surround was changed. However, when the surround was gray and black [shown by circles, input = (51.3, 41.85, 29.7) and (0, 0, 0)] there was a significant difference in output for different surround spatial structures.

attenuation when the high spatial frequencies in the inducing surround were due to luminance changes.

This behavior is also shown by the current network. Although the V4 stage linearly sums, its inputs from spectrally opponent cells, those inputs depend non-linearly on the spatial frequency properties of color and luminance variations in the image. High spatial frequencies in color cause an attenuation of the color signal at the spectrally opponent layer, while high spatial frequencies in luminance are enhanced. Therefore, if the color regions within the inducing surround are equiluminant, the presence of high spatial frequencies will attenuate the input to the spectrally specific contrast stage and subsequently will reduce the amount of induction relative to that induced by a homogeneous surround. To test this, we used an input image similar to that used by Wesner and Shevell (1990), a yellow test spot with either a red surround or a green surround. The red surround was either spatially homogeneous, or contained a thin ring around the test spot which was either black or a white which was equiluminant with the surround.

The results are shown in Fig. 15. The presence of the thin white ring in the surround significantly diminishes the color induction effect on the yellow center. However, the thin black ring, causes much less decrease in the color

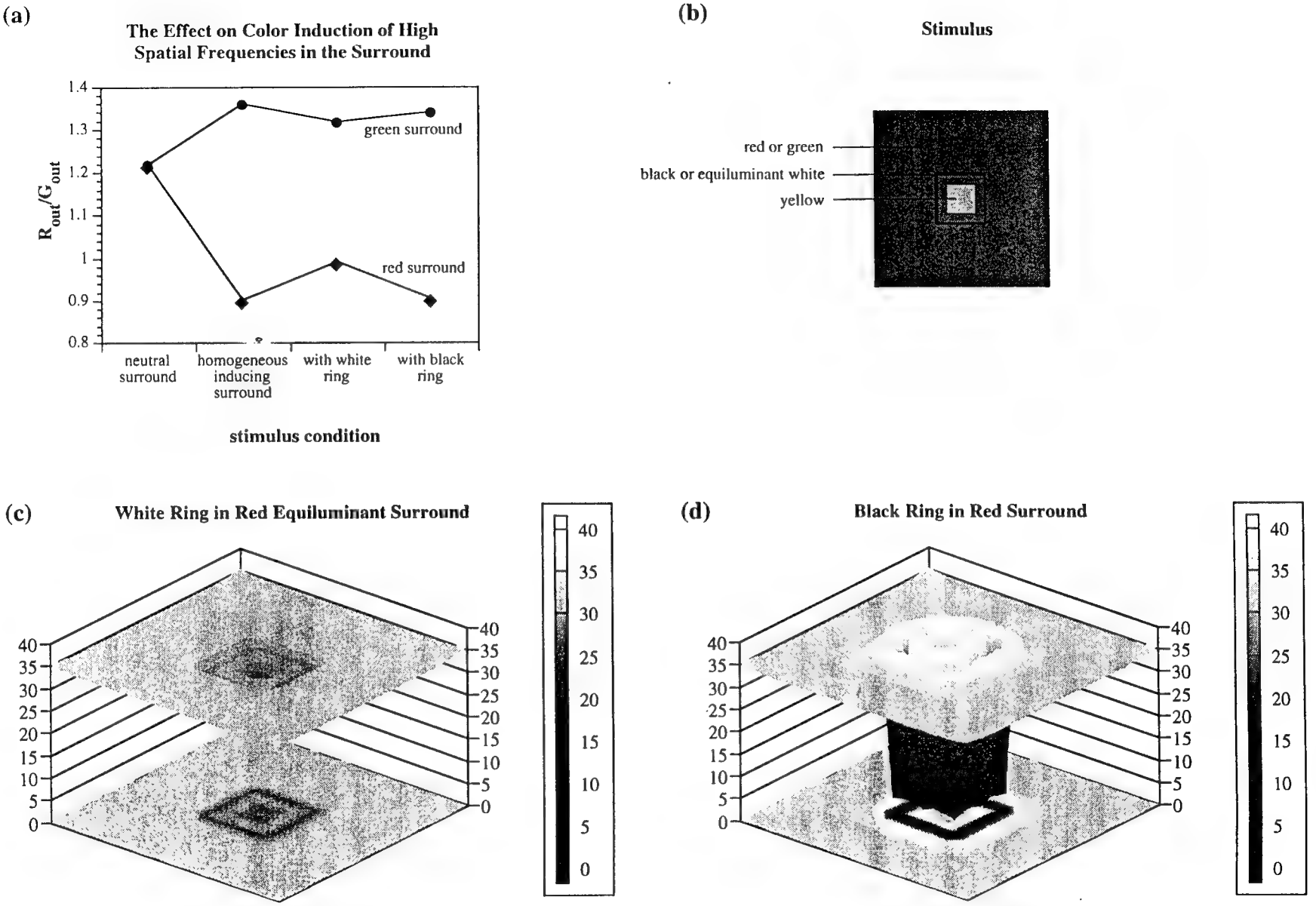


FIGURE 15. (a) Final output of the network for a yellow center stimulus with either a red or green surround. The surrounds were either homogeneous, contained a thin black ring, or contained a thin white ring which was equiluminant with the surround. (b) Spatial configuration of the stimulus. (c) Response of the R on-center spectrally and spatially opponent layer of the network to the yellow center, red background stimulus which includes a thin white ring around the center which is equiluminant with the red surround. (d) Response of the R on-center spectrally and spatially opponent layer of the network to the yellow center, red background stimulus which includes a black ring in the surround. The network shows an enhanced response to the red surround when the black ring is present, while the white ring induces no such response. The enhanced response around the black ring provides more input to the induction mechanisms of the V4 layer which compensates for the loss of input from the introduction of the black ring. Therefore, the white ring decreases the induction effectiveness of the red surround, but the black ring does not.

induction. The equivalent surround hypothesis (Valberg & Lange-Malecki, 1990) implies that the black ring would cause a greater decrease in induction than the white ring. The chromaticities of spatially averaged surrounds in the two cases are identical, but the luminance of the surround with the white ring is higher. Psychophysical experiments have shown that, in general, higher luminance surrounds have greater induction effects than lower luminance surrounds (Jameson, Hurvich & Varner 1979). However, the presence of the spectrally opponent stage before the spatial integration and induction of the spectrally specific contrast stage, results in the opposite effect. The luminance contrast of the black ring causes an enhancement of the response to the red surround in the area surrounding the black ring (see Fig. 15). The chromatic contrast of the white ring, on the other hand, causes a decrease in the response of the spectrally opponent cells to the red surround near the white ring.

The frequency dependent responses of the spectrally opponent cells provide the inputs to the final stage of the network. Therefore, the surround is spatially integrated only after the stimuli have been altered by the spectrally opponent stage. The net effect of the increased contrast response at the opponent stage in both color and luminance is to enhance the contribution of the spectrally-specific contrast stage which uses that contrast information in the push–pull mechanism. These enhancements are most significant when the positive and negative changes in inputs occur in cells which are operating in different parts of their response curves, and which, therefore, have different gains. In these cases, the changes caused by the opponent cells on either side of an edge do not cancel each other when the V4 cells sum these inputs. Because the opponent stage increases contrast under certain conditions, it can sometimes contribute to color and brightness induction via the V4 stage of the network. However, color induction decreases with increasing distance between center and inducing surround (e.g. Tiplitz-Blackwell & Buchsbaum, 1988a). The color opponent stage works in the opposite direction of these observations for high spatial frequency stimuli by decreasing color contrast. This property, combined with the lower sensitivity, compared to other stages, to global luminance changes and to spectrally-specific contrast, preclude consideration of the color opponent stage as a direct contributor to color constancy and color induction.

(iv) Nonlinearities

With limited adaptation, high or low luminance inputs can cause the cells to respond in the nonlinear portion of the response curve, saturating or rectifying respectively the cells' responses (see Figs 2 and 3). Allowing time for adaptation to take place prevents this to some degree, but if the luminance change is too large even long term adaptation will not compensate enough. The predicted color matches then move toward neutral (Fig. 16). This is analogous to surfaces appearing either "washed out" in very bright light or "muddied" in very dim light. However, this situation makes matching difficult because very different inputs (stimuli) can give very similar outputs (percepts).

The luminance levels of the images used for testing this system and the values for the slope and threshold of each cell type were chosen to prevent the cell responses from being significantly saturated or rectified under most circumstances. The nonlinearities were nonetheless important to the behavior of the network, however, because even at mid-range luminance levels saturated colors can cause individual color channels to respond outside of their linear range. The effect is that the gain of the induction shifts in each color channel will be different for different input images. For example, the gain of the induction in the B cells of the network is much less for a saturated yellow input than for a blue input. The effects of the nonlinearities on the spectrally-opponent stage responses were described in the previous section. However, the qualitative effects of the nonlinearities on the general behavior of the network were usually small.

DISCUSSION

We have described a model of color constancy and color induction which is based on aspects of the anatomy and physiology of the primate visual system. In a number of simulations, the model demonstrates a degree of constancy and induction which is similar to that shown by human psychophysics in previously reported experiments. The color constancy literature has generally been a retinal vs cortical mechanism debate. Although many have argued for a two-stage process (e.g. Walraven, 1976; Werner & Walraven, 1982; Arend & Reeves, 1986; Shevell *et al.*, 1992), the emphasis has often been on determining which stage is most important. The results of this simulation emphasize the importance of both retinal and cortical mechanisms in a cooperative multi-stage system. By systematically adding or eliminating each processing stage during the simulations, we could draw several conclusions about the contributions and interdependence of these stages.

We have shown that a simple push–pull contrast mechanism which uses V4 silent surround cells can, alone, accomplish a significant amount of both color constancy and color induction effects. Given some amount of spatio-temporal integration, adaptation alone can also achieve a degree of color constancy. We found that the effects of receptor adaptation and the V4 mechanism both depend on the particular stimulus conditions and that when one stage cannot achieve color constancy another stage can often compensate. Therefore, a system which combines both adaptation and cortical processing can achieve color constancy in a greater range of stimulus conditions than can either stage alone.

The approach taken here is quite different from many of the previous approaches to color constancy. Recently, much attention has been given to computational theories

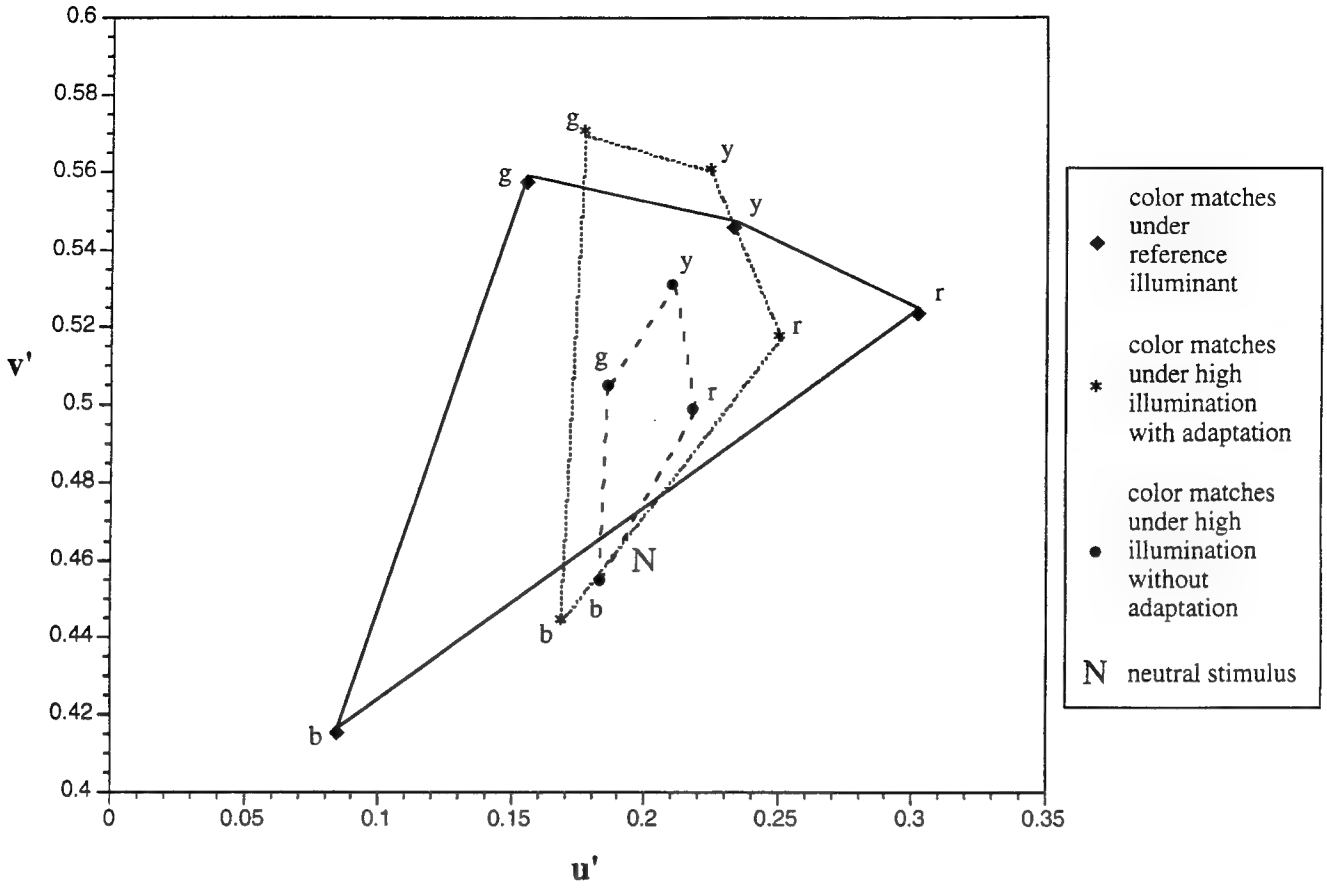


FIGURE 16. Four Munsell reflectances (10B6/10, 10GY7/10, 7.5Y7/10, 2.5YR7/10) were input to the network under a spectrally flat illuminant. They are plotted here in CIE $u'v'$ space. The "N" marks the chromaticity of the Munsell gray reflectance, N6.75, under the same illuminant. The squares and solid lines mark the matches, with all network stages active, under an illuminant whose luminance resulted in most of the cells in the simulation operating within their linear range. Matches to all colors under this reference illuminant had luminances between 30 and 43 cd m^{-2}. The triangles and dotted lines show the matches to the same reflectances under an illuminant 10 times greater. Even with adaptation, there is some shift of the matches to less saturated (i.e. closer to gray) colors. Without adaptation, circles and dashed lines, most cells' responses reach their maximum limit and the matches are very close to neutral. This demonstrates one effect of the nonlinear response function and the need for adaptation under changing luminance conditions.

for color constancy which estimate the surface reflectance by mathematically separating the illuminant from the reflectance. This is generally done by describing the reflectance and illuminant each as a sum of three basis functions (see review by Lennie & D'Zmura, 1988). The resulting set of equations is underdetermined. In order to solve this set of equations, these models require either restrictions on the reflectances, such as a gray average chromaticity, some a priori knowledge of the illuminant, or assumptions about the mathematical structure of reflectances and illuminants (Buchsbaum, 1978, 1980; Brill, 1978; Maloney & Wandell, 1986; D'Zmura & Lennie, 1986; Gershon & Jepson, 1989; Rubin & Richards, 1982; Dannemiller, 1989; Troost & de Weert, 1991a; Brainard & Wandell, 1991). A comprehensive mathematical analysis of the problem, generalizing the earlier approaches and using multiple surfaces and/or illuminants, is given by D'Zmura and Iverson (1993a, b).

There are additional restrictions which allow solutions to the reflectance–illuminant separation problem. One solution is to require the number of photoreceptors to be greater than the dimension of the reflectance space: (Maloney & Wandell, 1986). This solution enables simple reflectance–illuminant separation algorithms. However, this implies that either one must severely limit the reflectances that the algorithm can use or that more than three photoreceptor types be involved. Another attempt was made by Faugeras (1979) who developed a filter to separate the reflectance and illuminant by taking the logarithm of the reflectance–illuminant product, thus turning the product into a sum which may be separated by a linear filter. However, the algorithm encounters difficulties when both the illuminant and the reflectance vary so that their spatial Fourier spectra overlap. The current approach does not try to explicitly calculate the reflectance or illuminant spectrum, and so does not require any of these assumptions or restrictions.

Several "lightness" algorithms which have been previously proposed for color constancy, including the Retinex (Land & McCann, 1971), have been shown to be mathematically equivalent to a local spatial derivative plus a normalization term (Hurlbert, 1986). Similarly,

some have argued that the Retinex is essentially the same as Von Kries adaptation in that each is a renormalization of color channel activities relative to some white reference (see review by Jameson & Hurvich, 1989). In this sense, the adaptation stage and the spectrally-specific contrast stage in the simulation are also similar, as are the multiplicative adaptation and the subtractive adaptation mechanisms for color constancy described by previous researchers (Hayhoe *et al.*, 1987; Hayhoe & Wenderoth, 1991). However, there are several important differences in the operations described here which allow them to cover different stimulus conditions and, therefore, to cooperate in their contributions to color constancy.

First, the cone adaptation stage has a permanent white reference which is set by the midpoint of the range of possible threshold values. The adaptation stage also has a long-term adaptation reference which is usually close to neutral because it is established through exposure to many different stimuli over a long period of time. Faster, more localized adaptation effects are deviations from this long term reference. The reference for the spectrally-specific contrast stage is the activity of the "local reference" cells. The spectrally-specific contrast reference is not fixed, but instead changes with each new image. This reference is not usually neutral. The spectrally-specific contrast mechanism described here is also different from most "lightness" algorithms in that the normalizing reference is measured locally, rather than globally. In addition, the spatial profiles of the two constancy mechanisms are different. The effect of localized adaptation depends more heavily on the central test spot, while the effect of the large surrounds in the spectrally-specific contrast operation are more affected by the background stimulus.

The effect of the spectrally-specific contrast operation can be increased by the spectral and spatial opponency of the preceding stage which enhances color and brightness contrast for low spatial frequencies. The simulation results regarding the effects of image spatial structure on color induction are in agreement with the psychophysical results of Wesner and Shevell (1990). The simulation results confirm the assertion of Zaidi *et al.* (1992) that the psychophysical results could be explained by a mechanism which selectively attenuates high frequency chromatic stimuli before color induction takes place. The spectrally opponent cells reduce the effectiveness of high frequency chromatic inputs in the surround prior to spatial integration and induction by the spectrally-specific mechanism. This agreement of the model with the psychophysical data lends support to the idea that at least part of the color induction mechanism must lie beyond the stage which gives a low-pass response for color stimuli and a band-pass response for luminance stimuli. In other words, there are color induction mechanisms beyond the retina. If there are additional post-retinal contrast enhancing processes, these will also alter the equivalent surround of a complex image if these processes take place before the spatial integration in V4.

Another difference between the retinal and cortical stages which has not yet been incorporated into the current model is the existence of more than three distinct color channels in the cortex. This paper addresses the processing of color information in terms of color constancy and color induction, but does not address the more complex problems of image representation. Spectral sensitivities of cortical cells have been shown to have peaks at many different wavelengths (Zeki, 1980; Lennie, Krauskopf & Sclar, 1990; Schein & Desimone, 1990) indicating a more distributed representation of color information in the cortex. There are also questions remaining about how the processed color information is then integrated with information about image segmentation and object perception. The output of the simulation, which represents color information at a single point in the image is, most likely, highly simplistic.

Our goal was to examine the effects and interactions of color processing mechanisms rather than specific cellular mechanisms. We cannot rule out other possible implementations of the processing stages used in this network because many anatomical substrates can accomplish very similar processing tasks. We intentionally abstracted some of the anatomical details so that the emphasis would be on the information processing mechanisms themselves. The model is robust enough that the primary results do not depend on any particular parameter value or anatomical implementation. Eventually, we would like to make the simulation and our predictions for psychophysical and physiological experiments more quantitative. This will require more anatomical detail and more indepth parameter optimization. There are many parameters in the simulation which are not directly determinable from current physiological data. As a first step, however, we wished to address more general questions regarding color information processing in the visual system, independent of the specific anatomical implementation.

The implementation of the adaptation stage in the simulation was particularly difficult because it is a dynamic process in an otherwise static model. In order to calculate what the adaptation state should be, assumptions had to be made regarding the previous stimuli presented to each cone during the adaptation period. This depends on the conditions of the experiment being simulated. The results will be different for different types of viewing conditions (e.g. haploscopic, simultaneous match and test stimuli, or memory matches). We assumed in these simulations that prior to each stimulus presentation, there was long-term adaptation to a moderate luminance neutral uniform field. We calculated the adaptation shift (away from the neutral adaptation state) each time a new stimulus was presented, whether that stimulus was the test stimulus or the match. A gaussian weighting function was used because each image had a central region of interest and was either symmetrical about that central region, or had a Mondrian background which had a random distribution of color patches. As has been discussed in numerous psychophysical studies, the experimental conditions can

greatly affect the adaptation state of the visual system. The same is true with our simulation.

One possible extension of this model involves solving the problem of image scale invariance which, in the case of human color perception, means that the color of an object does not change significantly with size, provided that the regions surrounding that object are scaled in the same proportion. In other words, this is the common observation that objects don't change color as we walk toward them. One possible solution is dynamic receptive fields which adapt to match the spatial scale of the stimulus. Pettet and Gilbert (1992) have recently found physiological evidence for stimulus-dependent dynamic receptive fields in cortical area V1. In addition, Moran and Desimone (1985) reported cells in V4 and inferior temporal cortex whose responses depended on the state of attention of the animal. While evidence for very rapid stimulus dependent receptive field changes is still preliminary, if such mechanisms do exist then these dynamic properties could be incorporated into the V4 mechanism described here to allow for image scale invariance.

There are also additions that could be made in order to include other aspects of color perception. For example, it is known from psychophysics that there are contributions to brightness induction from binocular depth information (Schirillo & Shevell, 1993) and surface segmentation (White, 1979). There are also task dependent surface/illuminant segregation influences (Arend & Reeves, 1986; Troost & de Weert, 1991b; Craven & Foster, 1992). There are many interacting processes involved in color perception and no single mechanism can be credited with achieving "color constancy". In addition to color constancy and color induction, the stages of the network described here are rather basic processes and each stage is likely to serve many other roles in the visual system as well. The present study provides a different perspective in the debate as to whether retinal or cortical mechanisms have a greater contribution to color constancy and color induction. Although others have suggested the need for both retinal and cortical visual color processing, this paper emphasized the distinct roles of each stage and the interactions between the stages. The two levels of processing have important but different effects on color constancy and color induction, not necessarily greater or smaller effects.

REFERENCES

Arend, L. & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America A, 3*, 1743–1751.

Beck, J. (1972). *Surface color perception*. London: Cornell University Press.

Berman, N. J., Douglas, R. J., Martin, K. A. & Whitteridge, D. (1991). Mechanisms of inhibition in cat visual cortex. *Journal of Physiology, London, 440*, 697–722.

Brainard, D. H. & Wandell, B. (1991). A bilinear model of the illuminant's effect on color appearance. In Movshon, J. A. & Landy, M. S. (Eds), *Computational models of visual processing*. MIT Press: Cambridge, Mass.: MIT Press.

Brainard, D. H. & Wandell, B. (1992). Asymmetric color matching: How color appearance depends on the illuminant. *Journal of the Optical Society of America A, 9*, 1433–1448.

Brill, M. H. (1978). A device performing illuminant-invariant assessment of chromatic relations. *Journal of Theoretical Biology, 71*, 473–478.

Brill, M. H. & West, G. (1986). Chromatic adaptation and color constancy: A possible dichotomy. *COLOR Research and Application, 11*, 196–204.

Brown, R. O. (1993). Integration of enhanced-contrast edges in color vision. *Investigative Ophthalmology & Visual Science, 34*, 766.

Buchsbaum, G. (1978). Models of central signal processing in colour perception. Dissertation, Tel-Aviv University, Israel.

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin Institute, 310*, 1–26.

Cornellissen, F. W. & Brenner, E. (1991). On the role and nature of adaptation in chromatic induction. In Blum, B. (Ed.), *Channels in the visual nervous system: Neurophysiology, psychophysics and models* (pp. 109–124). Tel Aviv: Freund.

Craven, B. J. & Foster, D. H. (1992). An operational approach to colour constancy. *Vision Research, 32*, 1359–1366.

Dannemiller, J. L. (1989). Computational approaches to color constancy: Adaptive and ontogenetic considerations. *Psychological Review, 96*, 255–266.

Derrington, A. M. & Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *Journal of Physiology, 357*, 219–240.

Desimone, R. & Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: Sensitivity to stimulus form. *Journal of Neurophysiology, 57*, 835–868.

Desimone, R., Moran, J., Schein, S. J. & Mishkin, M. (1993). A role for the corpus callosum in visual area V4 of the macaque. *Visual Neuroscience, 10*, 159–171.

Dufort, P. A. & Lumsden, C. J. (1991). Color categorization and color constancy in a neural network model of V4. *Biological Cybernetics, 65*, 293–303.

D'Zmura, M. & Iverson, B. (1993a). Color constancy. I. Basic theory of two-stage linear recovery of spectral descriptions for lights and surfaces. *Journal of the Optical Society of America A, 10*, 2148–2165.

D'Zmura, M. & Iverson, G. (1993b). Color constancy. II. Results for two-stage linear recovery of spectral descriptions for lights and surfaces. *Journal of the Optical Society of America A, 10*, 2166–2180.

D'Zmura, M. & Lennie, P. (1986). Mechanisms of color constancy. *Journal of the Optical Society of America A, 3*, 1662–1672.

Faugeras, O. D. (1979). Digital color image processing within the framework of a human visual model. *IEEE Transactions on Acoustics, Speech and Signal Processing, 27*, 380–393.

Gershon, R. & Jepson, A. D. (1989). The computation of color constant descriptors in chromatic images. *COLOR Research and Application, 14*, 325–334.

Grossberg, S. (1987). Cortical dynamics of three-dimensional form, color and brightness perception. *Perception & Psychophysics, 41*, 87–158.

Hayhoe, M. M. & Wenderoth, P. (1991). Adaption mechanisms in color and brightness. In Valberg, A. & Lee, B. (Eds), *From pigments to perception* (pp. 353–367). New York: Plenum Press.

Hayhoe, M. M., Benimoff, N. I. & Hood, D. C. (1987). The time-course of multiplicative and subtractive adaptation process. *Vision Research, 27*, 1981.

von Helmholtz, H. (1866). *Physiological optics* (Trans. Southall, J. P. C., 1924 (3rd edn, pp. 286–287). Rochester, N.Y.: Optical Society of America.

Helson, H. (1938). Fundamental problems in color vision. I: The principle governing changes in hue, saturation, and lightness of non-selective samples in chromatic illumination. *Journal of Experimental Psychology, 23*, 439–476.

Hering, E. (1878). A theory of the light sense (Trans. Hurvich, L. & Jameson, D.). Cambridge, Mass.: Harvard University Press.

Heywood, C. A., Gadotti, A. & Cowey, A. (1992). Cortical area V4 and its role in the perception of color. *Journal of Neuroscience, 12*, 4056–4065.

Hurlbert, A. (1986). Formal connections between lightness algorithms *Journal of the Optical Society A, 3*, 1684–1693.

Hurlbert, A. C. & Poggio, T. A. (1988). Synthesizing a color algorithm from examples. *Science, 239*, 482–485.

Jameson, D & Hurvich, L. M. (1989). Essay concerning color constancy. *Annual Review of Psychology, 40*, 1–22.

Jameson, D., Hurvich, L. M. & Varner, F. D. (1979). Receptoral and postreceptoral processes in recovery from chromatic adaptation. *Proceedings of the National Academy of Sciences, U.S.A., 76*, 3034–3038.

Judd, D. B. (1940). Hue, saturation, and lightness of surface colors with chromatic illumination. *Journal of the Optical Society of America, 30*, 2–32.

Land, E. H. & McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America, 61*, 1–11.

Land, E. H., Hubel, D. H., Livingstone, M. S., Perry, S. H. & Burns, M. M. (1983). Colour-generating interactions across the corpus callosum. *Nature, 303*, 616–618.

Lennie, P. & D'Zmura, M. (1988). Mechanisms of color vision. *Critical Reviews in Neurobiology, 3*, 333–401.

Lennie, P., Krauskopf, J. & Sclar, G. (1990). Chromatic mechanisms in striate cortex of macaque. *Journal of Neuroscience, 10*, 649–669.

Lucassen, M. & Walraven, J. (1993). Quantifying color constancy: Evidence for nonlinear processing of cone-specific contrast. *Vision Research, 33*, 739–757.

Maloney, L. T. & Wandell, B. A. (1986). Color constancy: A method for recovering surface spectral reflectance. *Journal of the Optical Society of America, 3*, 29–33.

McCann, J. J., McKee, S. P. & Taylor, T. H. (1976). Quantitive studies in Retinex theory: A comparison between theoretical predictions and observer responses to the 'Color Mondrain' experiments. *Vision Research, 16*, 445–458.

Moore, A., Allman, J. & Goodman, R. M. (1991). A real-time neural system for color constancy. *IEEE Transactions on Neural Networks, 2*, 237–246.

Moran, J. & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*, 782–784.

Moran, J., Desimone, R., Schein, S. J. & Mishkin, M. (1983). Suppression from ipsilateral visual field in area V4 of the macaque. *Society for Neuroscience Abstracts, 9*, 957.

Naka, K. & Rushton, W. A. H. (1966). S-potentials from luminosity units in the retina of fish (Cyprinidiae). *Journal of Physiology, 188*, 587–599.

Pettet, M. W. & Gilbert, C. D. (1992). Dynamic changes in receptive-field size in cat primary visual cortex. *Proceedings of the National Academy of Sciences, 89*, 8366–8370.

Rubin, J. M. & Richards, W. A. (1982). Color vision and image intensities: When are changes material? *Biological Cybernetics, 45*, 215–226.

Sajda, P. & Finkel, L. H. (1992). NEXUS: A simulation environment for large-scale neural systems. *Simulation, 59*, 358–364.

Schein, S. J. & Desimone, R. (1990). Spectral properties of V4 neurons in the macaque. *Journal of Neuroscience, 10*, 3370–3389.

Shevell, S. K. & Wesner, M. F. (1990). Chromatic adapting effect of an achromatic light. *Optical Society of America Technical Digest Series, 11*, 104.

Shevell, S. K., Holliday, I. & Whittle, P. (1992). Two separate neural mechanisms of brightness induction. *Vision Research, 32*, 2331–2340.

Schirillo, J. A. & Shevell, S. K. (1993). Lightness and brightness judgements of coplanar retinally noncontiguous surfaces. *Journal of the Optical Society of America A, 10*, 2442–2452.

Tiplitz Blackwell, K. & Buchsbaum, G. (1988a). The effect of spatial and chromatic parameters on chromatic induction. *COLOR Research and Application, 13*, 166–173.

Tiplitz Blackwell, K. & Buchsbaum, G. (1988b). Quantitive studies of color constancy. *Journal of the Optical Society of America, 5*, 1772–1780.

Troost, J. M. & de Weert, C. M. M. (1991a) Surface reflectances and human color constancy: Comment on Dannemiller (1989). *Psychological Review, 98*, 143–145.

Troost, J. M. & de Weert, C. M. M. (1991b). Naming versus matching in color constancy. *Perception & Psychophysics, 50*, 591–602.

Valberg, A. & Lange-Malecki, B. (1990). 'Color constancy' in Mondrian patterns: A partial cancellation of physical chromaticity shifts by simultaneous contrast. *Vision Research, 30*, 371–380.

Von Kries, J. (1905). Die Gesichtsempfindungen. In Nagel, W. (Ed.), *Handbuch der Physiologie der Menschen* (pp. 109–282). Brunswick, N.J.: Vieweg.

Vos, J. J., (1978). Colorimetric and photometric properties of a 2° fundamental observer. *COLOR Research and Application, 3*, 125–128.

Vos, J. J. & Walraven, P. L. (1971). On the derivation of the foveal receptor primaries. *Vision Research, 11*, 799–818.

Walraven, J. (1976). Discounting the background, the missing link in the explanation of chromatic induction. *Vision Research, 16*, 289–295.

Walraven, J. & Werner, J. S. (1991). The invariance of unique white: Possible implications for normalizing cone action spectra. *Vision Research, 31*, 2185–2193.

Walraven, J., Enroth-Cugell, C., Hood, D. C., MacLeod, D. I. A. & Schnapf, J. L. (1990). The control of visual sensitivity: Receptoral and postreceptoral processes. In Spillman, L. & Werner, J. S. (Eds), *Visual perception: The neurophysiological foundations*. San Diego, Calif.: Academic Press.

Walsh, V., Kulikowski, S. R., Butler, S. R. & Carden, D. (1992). The effects of lesions of area V4 on the visual abilities of macaques: Color categorization. *Behavioral Brain Research, 52*, 82–89.

Werner, J. S. & Walraven, J. (1982). Effect of chromatic adaptation on the chromatic locus: The role of contrast, luminance, and background color. *Vision Research, 22*, 929–943.

Wesner, M. F. & Shevell, S. K. (1992). Color perception within a chromatic context: Changes in red/green equilibria caused by non-contiguous light. *Vision Research, 32*, 1623–1634.

Wesner, M. F. & Shevell, S. K. (1993). Color appearance and chromatic surrounds. *Investigative Ophthalmology & Visual Science, 34*, 746.

White, M. (1979). A new effect of pattern on perceived lightness. *Perception, 8*, 413–416.

Wyszecki, G. & Stiles, W. S. (1982). *Color science* (2nd edn). New York: Wiley.

Yoshioka, T., Levitt, J. B. & Lund, J. S. (1992). Intrinsic lattice connections of macaque monkey visual cortex. *Journal of Neuroscience, 12*, 2785–2802.

Zaidi, Q., Yoshimi, B., Flanigan, N. & Canova, A. (1992). Lateral interactions within color mechanisms in simultaneous induced contrast. *Vision Research, 32*, 1695–1708.

Zeki, S. (1980). The representation of colors in the cerebral cortex. *Nature, 284*, 412–418.

Zeki, S. M. (1983). Color coding in the cerebral cortex: The reaction of cells in monkey visual cortex to wavelengths and colors. *Neuroscience, 9*, 741–765.

Acknowledgements—The authors would like to thank Paul Sajda for extensive help in implementing the network using the NEXUS neural simulator. This work was supported by grants from the Office of Naval Research, N00014-90-J-1864 and N00014-93-1-0681; Air Force Office of Scientific Research, 91-0082 and 92-J-0316, The Whitaker Foundation, and the McDonnell-Pew Program in Cognitive Neuroscience.

A Multi-Stage Neural Network for Color Constancy and Color Induction

Susan M. Courtney, Leif H. Finkel, Gershon Buchsbaum
Department of Bioengineering, and Institute for Neurological Sciences,
University of Pennsylvania,
220 South 33rd Street, Philadelphia, PA 19104

IEEE Transactions on Neural Networks (in press)

A biologically-based multi-stage neural network is presented which produces color constant responses to a variety of color stimuli. The network takes advantage of several mechanisms in the human visual system, including retinal adaptation, spectral opponency, and spectrally-specific long-range inhibition. This last stage is a novel mechanism based on cells which have been described in cortical area V4. All stages include non-linear response functions. The model emulates human performance in several psychophysical paradigms designed to test color constancy and color induction. We measured the amount of constancy achieved with both natural and artificial simulated illuminants, using homogeneous grey backgrounds and more complex backgrounds, such as Mondrians. On average, the model performs as well or better than the average human color constancy performance under similar conditions. The network simulation also displays color induction and assimilation behavior consistent with human perceptual data.

INTRODUCTION

Color perception can contribute to object recognition if the perceived color is a fixed attribute of the object surface. However, the color signal which reaches the eye, or any other detector, is the product of the surface reflectance and the incident illuminant. Therefore, this signal changes as the illuminant changes, from noon sunlight to hazy sunset, or from incandescent to fluorescent artificial lighting. The term "color constancy" has been used to describe the ability of humans and other animals [25, 30] to discount a portion of the illuminant in order to make more reliable judgments of surface color. Color induction, a related phenomenon, is the change in the perceived color of a surface due to its juxtaposition with other colored surfaces. Color induction enhances the color contrast in a scene and probably aids in object detection and surface segmentation. Color constancy and color induction demonstrate that human color perception depends on the spatial distribution of the wavelengths of light present in the entire image. Cameras and most artificial vision systems, on the other hand, are not color constant. Photographs taken in fluorescent light often look green.

We would like both to understand how the human visual system accomplishes the task of discounting the illuminant, and also to build a practical artificial vision system which has color constancy. These are not necessarily distinct tasks. A logical way to design an artificial color constant system is by reverse engineering of the existing biological design, the human visual system. An ideal color constant system should show the following characteristics:

1) Accurate object reflectance color determination

2) Flexibility, *i.e.* good color output for a large variety of images.

3) Few assumptions about the image and little *a priori* knowledge.

The human visual system, as will be explained in what follows, seems to favor flexibility and fewer assumptions over complete accuracy. The color constancy achieved by the biological system is thus only approximate, but it is able to handle a large variety of scenes under very different lighting conditions. Depending upon the application, one may want to design a system which has perfect color constancy, one which more closely mimics the human visual system, or one which is a compromise between the advantages and disadvantages of each. An example of an application which would require more of an emphasis on accurate color discrimination would be quality control of textiles in multiple factories which have different lighting conditions. An application which would emphasize the duplication of human color perception would be catalogue printing in which the color of a picture of an item as perceived by a human looking at the catalogue should be the same as the color perceived by a human looking at the actual object being sold.

Most previous computational algorithms for color constancy [3, 8, 13, 14, 15, 17, 37, 42] have favored accuracy over flexibility. Most of these try to explicitly recover the reflectance spectrum from the reflectance illuminant product. Because the resulting set of equations is underdetermined, assumptions regarding either the reflectances or the illuminant must be made in order to find a solution. When these assumptions are not met, the algorithms can make color predictions far from those reported by humans.

One algorithm for achieving color constancy which has received much attention is the retinex theory developed by Land [31-34]. The retinex calculates the relative "lightness" of each area of a scene within three separate channels, each sensitive to a different region of the visual spectrum. A variety of algorithms have been suggested for computing the lightness values. (For example [31-34], [38], [23]) The most recent method of calculation [33] bears a resemblance to the receptive fields of cells recorded by Schein and Desimone [44] in cortical area V4 in that it effectively subtracts a large "surround" region from a very small "center" with similar wavelength sensitivity.

The retinex theory successfully predicts many of the basic properties of human color perception, but there are limitations in the algorithm in certain situations as has been demonstrated by Brainard and Wandell [2] and in a neural network by Moore, Allman, and Goodman [39]. Because the retinex records only contrast, the interior of large uniformly colored areas become grey. In addition, in order for the contrast information to be converted to color and luminance information, one of two assumptions must be made. Either the average chromaticity of the reflectances in a scene must be constant for all

images (known as the grey world assumption) or the brightest region in an image must be a white reflectance. In addition, the retinex must assume that the illumination varies slowly in space while reflectances have sharp chromatic borders. Some of these assumptions were also used by many of the separation of reflectance and illuminance theories. There are many examples of images that do not fit these conditions (such as scenes with shadows or simple images which contain a strongly colored object). In such cases the retinex will predict colors far from those perceived by humans.

Neural networks have been used for implementing the Retinex and a variety of other color constancy algorithms. As mentioned above, Moore et al. [39] implemented Land's retinex algorithm within a neural network structure. They also incorporated a modification in order to eliminate the color washout problem described above by multiplying the surround factor by the "edginess" of the area. Hurlbert and Poggio [24] demonstrated that, using a number of different "learning" methods in a neural network, including least squares gradient descent and back propagation, a linear operator could be found which is similar to that proposed in the Retinex algorithm.

Grossberg [18] also developed a neural network simulation which obtains color information from the contrast at boundaries and then fills in the color into each segment of the image. It successfully demonstrates many aspects of brightness and color perception. However, because only the contrast information is preserved, the system must rely on the "grey world" assumption to determine the mean level to which the contrast refers. Therefore, as with all theories which use this assumption, the predictions of the network can be inconsistent with psychophysical results in scenes that have a non-neutral average chromaticity. Also, because it fills in each segment with the average color determined from the surrounding boundaries, it loses information about subtle variations in the reflectance within a segment.

The network model developed by Dufort and Lumsden [12] uses double opponent cells and features output cells which behave qualitatively like the "color constant" cells reported by Zeki [58] in V4. The parameters of the network were optimized to create output response curves which correspond to color naming categories developed by psychophysical studies. The network design directly incorporates several aspects of color psychophysics. At present, however, the network addresses only hue constancy. The saturation and lightness dimensions of color signals are also affected by changes in the illuminant, and require a constancy mechanism as well.

This paper describes a system which is based on the primate visual system, from the retinal to cortical area V4. Each stage contributes both to producing better color constancy and to the flexibility of the system. No assumptions are made about the

mathematical structure of the reflectances and illuminants, the average chromaticity of the reflectances in the image, or about the existance of chromatic or luminance gradients in either the reflectances or illuminants.

NETWORK ARCHITECTURE

An overview of the processing mechanisms in the network is shown in figure 1. The network was simulated using NEXUS, an interactive neural simulator designed for large scale models [43]. The complete network consists of over 11,000 cells and approximately 1.65 million connections. Below we will describe the properties of each stage of the network. Table 1 summarizes the most significant parameters in the model.

i. Input

The first stage corresponds to the cone responses. The input image is a 27x27 array, in which each entry defines the color at that location, specified in either Munsell color notation or in CIE (Commission Internationale de l'Eclairage) notation (x, y, Y). The array is converted to three 27x27 arrays of cone activation levels: R,G,B. Therefore, an input image unit has a one-to-one correspondence with a set of three units (analogous to one cone of each type) in the first layer of the network. For example, an image defined using Munsell spectra would be converted, at each point, to the three normalized cone activation levels by using the Vos-Walraven [51, 52] cone action spectra $(r(\lambda),g(\lambda),b(\lambda))$ which are shown in figure 2.

$$
\begin{aligned}
R &= \sum_{\lambda=400}^{700} k_1 r(\lambda)\mathcal{R}(\lambda)I(\lambda)\Delta\lambda \\
G &= \sum_{\lambda=400}^{700} k_2 g(\lambda)\mathcal{R}(\lambda)I(\lambda)\Delta\lambda \qquad (1) \\
B &= \sum_{\lambda=400}^{700} k_3 b(\lambda)\mathcal{R}(\lambda)I(\lambda)\Delta\lambda
\end{aligned}
$$

where $\mathcal{R}(\lambda)$ is the reflectance spectrum and $I(\lambda)$ is the illuminant. The coefficients $k_{1,2,3}$ are constants which normalize the sensitivity spectra so that all cone types in the simulated array have the same peak sensitivity. Therefore, the three types of first layer units ("cones") have responses of the same order of magnitude, and we may use the same dynamic range for all chromatic cell types in subsequent stages. For those cases in which the image was specified in CIE notation, the image was converted to cone activation levels by applying the transformations for Vos-Walraven action spectra [51, 52, 56] and then normalized using the same coefficients, $k_{1,2,3}$:

$$R = k_1 Y\left[0.155\left(\frac{x'}{y'}\right) + 0.543 - 0.037\left(\frac{1-x'-y'}{y'}\right)\right]$$

$$G = k_2 Y\left[-0.155\left(\frac{x'}{y'}\right) + 0.457 - 0.030\left(\frac{1-x'-y'}{y'}\right)\right] \qquad (2)$$

$$B = k_3 Y\left[0.007\left(\frac{1-x'-y'}{y'}\right)\right]$$

where x' and y' are the Judd modified [24] 1931 CIE chromaticity coordinates. Y is the luminance in cd/m^2 of the stimulus and is used here to scale the cone responses for luminance.

ii. Cell Responses and Nonlinearities

In the simulation of the network model, the input layers of the network correspond to the three cone types of the human retina. These cells have a Naka-Rushton type response function [41]:

$$A_i = \frac{Q_i^x}{Q_i^x + \sigma_i^x} \qquad (3)$$

where Q_i is the total input to cell i, σ_i is the threshold of cell i, and x is a constant from 0.7 to 1.0. In the simulation results shown here x=0.9, but the general behavior of the system was not very sensitive to the value of this parameter. In all other stages, cell activity is determined by a standard sigmoidal response function of the input:

$$A_i = (max - min)\left(\frac{1}{1 + e^{-(Q_i - \sigma_i)\beta_i}}\right) + min. \qquad (4)$$

iii. Adaptation

(a) Motivation

The nonlinear response function of both the cones and the cells in higher layers of the network, gives the system a limited dynamic range. However, the light level varies in our daily environment from 10^{-4} to 10^{+5} cd/m^2. In order to keep the cell responses in all stages within the linear range of their response functions, the cones must be able to shift their thresholds to accomodate the overall level of incoming light. Although there may also be mechanisms for adaptation in later stages of the visual system, individual primate cones are known to change their sensitivities according to the amount of light available [46].

Adaptation was recognized early on as a probable contributor to color constancy in humans. (e.g., [21], [50]) Although adaptation alone, defined as a multiplicative gain

change of individual photoreceptors, has been mathematically proven to be incapable of achieving perfect color constancy [6, 11], it can produce some degree of color constancy if the mechanism includes integration across space and time. In human psychophysical experiments, color constancy has been shown to depend on length of presentation time [4, 19, 20] and eye movements [9]. This implicates receptor adaptation because it depends on the temporal integration of activity in spatially localized mechanisms. Therefore, the inclusion of adaptation in our simulation allows us to better mimic human perception by incorporating the differences in color perception under different viewing conditions.

(b) Implementation

In the simulation, we assume an initial long-term adaptation to a uniform neutral background. (see [53] for a review of psychophysical and physiological studies on adaptation) The amount of threshold shift is determined by the difference between the cone activation level for the neutral background stimulus and the cone activation level for the new stimulus. Because adaptation is dependent on the temporally weighted average of its input, the adaptation shift for a cone is dependent, not only on the point in the image directly corresponding to that cone position, but also on the surrounding area to which the cone may be exposed during eye movements, or from optical blur.

We approximated this temporal averaging effect by a two-dimensional Gaussian spatial weighting function, because for most viewing conditions which have been tested psychophysically, there is either a fixation point, or a central test patch around which one can assume eye movements were centered. In the simulation, the amount of the shift follows a sigmoidal function of the difference between the neutral and the current stimuli and is proportional to the length of viewing time. These constraints are incorporated into the simulation by including one adaptation cell for each receptor. The adaptation cell shifts the threshold of its corresponding receptor according to the following equation:

$$\sigma_{new_i} - \sigma_{neut} = \alpha\left\{2G\left(\frac{1}{1+e^{-(Q_i - Q_{neut})\beta_i}}\right) - G\right\}$$

$$G = \left|\sum_{i=0}^{n} (Q_i - Q_{neut})\left(\frac{1}{2\pi\theta^2}\right)e^{\left(\frac{-(x^2+y^2)}{2\theta^2}\right)}\right| \qquad (5)$$

where

σ = threshold

β = proportional to the slope of the linear portion of the function.

Q_i = cone activation level (i.e. R, G, or B)due to current image pixel *i*

Q_{neut} = cone activation level due to standard neutral input at image pixel i

n = total number of matrix entries comprising the image,

x , y = the horizontal and vertical distances from entry *i* to the center of the cone's receptive field when fixated on the center of the image

θ = the width of a Gaussian weighting function which varies with the degree of fixation required for the experiment

α = the fraction achieved within the stimulus presentation time of the total difference in long-term adaptation states between the neutral state and the state for the new stimulus.

α is proportional to the time of exposure. As α increases, the size of the threshold shift increases, following a sigmoidal curve ranging from -G to +G, where G is the difference between the weighted average activation level for the current image and the activation level for a uniform neutral background. (See figure 3) In the current study, α was held fixed at 0.3 and θ was held fixed at 3.0. However, we wished to include this flexibility in the model so that in future studies of the current network, we could make closer comparisons with human psychophysical data. With longer exposure time, the adapting cell will be able to better adapt (larger α) to its new stimulus. Under certain experimental conditions, longer exposure time may also allow for more eye movements. The spatial extent of the weighting function broadens (larger θ) with more eye movements. In the extreme case of very long exposure time and completely random eye movements over the entire field of view, the weighting function would be flat and the cone would adapt to the field average. This dependence of the parameters θ and α on eye movements and viewing time allows the effects of the adaptation stage of the simulation to vary with the type of viewing conditions being considered. (See Table 1 for additional information on parameter values.)

iv. Spectral opponency

Spectrally opponent cells are excited by one region of the spectrum and inhibited by a different region. In the network spectrally opponent cells are obtained by subtracting responses of spectrally opponent cone types and are generally based on the properties of LGN parvocellular type I receptive fields, which are spatially opponent as well. [22] This means that the excitatory and inhibitory regions are spatially segregated into center and surround regions. The spectral opponency gives the cell a high gain for low spatial frequency color changes, *i.e.* a color change which is spatially homogeneous across both

center and surround regions. The reason for this high gain is that such a color change will result in the combination of either an increase in excitation and a decrease in inhibition or a decrease in excitation and a increase in inhibition. For example, for a cell which is excited by red light in the center of its receptive field and inhibited by green light in its surround, a color change of the stimulus from green to red will result in increased excitatory input to the center and decreased inhibitory input to the surround.

High spatial frequency stimuli, such as edges, for which the center and surround regions receive different inputs, result in a different type of response. In the case of high spatial frequency stimuli which differ only in luminance, spatial opponency results in enhanced cell responses, because the excitatory regions may receive input from the higher luminance portion of the stimulus while the inhibitory regions receive input from the lower luminance portion of the stimulus, or vice versa. High spatial frequency stimuli which differ only in color, however, may result in diminished responses. For example, a red-green edge stimulating a red excitatory center, green inhibitory surround cell, will result in a smaller response than would a homogeneous red stimulus, because the green portion of the stimulus will inhibit the surround region of the receptive field, even though the center may be receiving excitation from the red side of the stimulus edge. This will result in the blurring of purely chromatic edges (i.e. edges with no luminance difference).

A layer of spectrally and spatially opponent cells contributes to the color constancy abilities of the current system in two ways. First, the increased chromatic gain enhances the overall sensitivity of the final stage of the network. Second, the additional enhancement of luminance edge responses acts in a way analogous to the "edginess" factor used by Moore *et al.* [39] to eliminate the washout of the color response to large homogeneously colored regions seen in the Retinex algorithm [33]. In their improved color constancy network, the main calculation was

$$\text{output} = \text{center} - (\text{surround} * \text{edginess}) \qquad (6)$$

where the edginess was determined by the average of the absolute values of the local spatial derivatives. The spatially opponent second stage of the current network accomplishes something similar in that the responses of this layer are enhanced in regions with luminance edges. One difference between the current operation and the Moore *et al* edginess factor is that the Moore *et al* . operation enhances both color and luminance edges. The spectrally and spatially opponent cells enhance luminance edge responses, but diminish responses at equiluminant color edges. In real images, however, equiluminant edges are rare. These are mostly laboratory color stimuli designed to control effects of luminance. The enhanced response provides the input to the final stage of the network, described in the next section, which is also a center-surround operation. Therefore, the

final stage will produce larger color shifts in "edgy" regions and smaller shifts in large homogeneously colored regions.

To produce these cells in the simulation, each "cell" receives excitatory input from a single cone in the center of its receptive field and inhibitory input from both R and G cone types surrounding the center [35]. No inhibitory surround in the network receives input from B cones because physiological recordings show no significant B cone input in LGN cell inhibitory surrounds [35, 60] The surround input is most heavily weighted toward the cone type(s) opponent to the center cone type. The number of R and G cells providing input to the surrounds are equal for all cell types. However, for R center cells, the amplitude of the synaptic weighting function for the G cone input to the surround is twice that of the R cone input to the same cell. The center strength (volume of 2D Gaussian sensitivity profile) is twice that of the surround, allowing these cells to have a significant response to homogeneous fields as well as to edges.

v. Higher Cortical Processing

(a) Motivation

The next stage in the network is similar to the Retinex and related color constancy algorithms in that it measures and uses spectrally-specific contrast [36, 48], what Land called "lightness" [31]. However, the method of normalizing contrast responses relative to a reference level is handled differently in the current system. In this simulation, the final stage is designed to respond according to the primary chromatic properties of cells in V4 [44]. Most cells in V4 have large, suppressive surrounds which have approximately the same wavelength sensitivity as the center of the receptive field. (See fig. 1) These large surrounds are called "silent surrounds" because they had little or no effect on the cell's activity unless the center was also stimulated. Desimone and Schein and their colleagues [10, 40] reported that the effect of stimulation in the silent surround decreases with increasing distance from the classical receptive field. Psychophysical results also show a decrease in the effect of inducing regions with distance. (e.g., [47], [49], [54], [57])

The strengths of the centers and silent surrounds of V4 cells appear to be well balanced; stimulation of the surround can completely inhibit the response to stimulation of the center [44]. Because the cells in V4 with silent surrounds respond only when there is a difference, either in wavelength or luminance, between the center and the distant surround, these cells are particularly well suited for carrying information about spectrally specific contrast. The significance of spectrally specific contrast in the visual system has been demonstrated psychophysically [36, 38, 48]. However, for those images that have little spectrally specific contrast, or an unknown or non-grey average chromaticity (e.g. blue

sky, green forest), the DC (or spatial average) information is also important. It is significant, therefore, that approximately 10% of the cells found in V4 did not have silent surrounds. The cells without silent surrounds have the same classical receptive field response as those cells with silent surrounds. These cells have the capacity to carry the (spatial) DC portion of the signal, *i.e.* to respond to homogeneous fields as well as edges and small spots. These center-only cells have been included in the network and we refer to them as "local reference cells" because in the network they provide the normalizing reference information for the contrast cell responses.

(b) Implementation

To incorporate these observations into the simulation, the responses of analogous V4 stage "cells" in the simulation were created directly using the outputs of the spectrally opponent stage. A positive contrast cell receives its input, excitatory from the center and inhibitory from the surround, from a single type of on-center spectrally opponent cell. Therefore, the positive contrast cells respond to images for which the input to its classical receptive field is greater than the input to its silent surround. We have also included negative contrast cells which receive input from off-center cells and, therefore, respond when the center input is less than the surround input. The inputs to the surround are weighted according to distance from the center by a negative exponential function. (See figure 1.) The "silent" nature of the V4 cell surrounds was implemented in the simulation by rectified inhibition, which was achieved by giving the V4 cells in the simulation, like those recorded physiologically, very low levels of spontaneous activity. Therefore, inhibitory input from the surround was only effective when there was also excitatory input from the center, "classical" receptive field.

In order to combine the physiological information from the local reference and contrast cells into a simple set of outputs which could be compared to human color perception, we combined the outputs of these V4-like cells into a simple push-pull mechanism. (This stage is shown in figure 1b.) This is the output of the final network stage and it is determined by the response of the local reference cells, enhanced by the positive contrast cells, or inhibited by the negative contrast cells, as given by the equation

$$O_x = B_x + c_1 P_x - c_2 N_x \qquad x = R, G, B \qquad (7)$$

where O is the output, B is the local reference response, P is the positive contrast response, N is the negative contrast response, and c_1 and c_2 are constants. The constants c_1 and c_2 are chosen, together with α (in equation 5), to increase or decrease the size of the constancy shift. In the simulation, we used $\alpha = 0.3$ and $c_1 = c_2 = 0.2$. This choice

enabled a minimum of 20% constancy (measured as distance in R G B space, see below) for all stimuli. In addition, these parameter values resulted in network behavior which was similar to the corresponding psychophysical data [48]

RESULTS

i. Measuring Constancy

An output to a particular reflectance in an image is considered as *"achieving some degree of color constancy"* if the difference (in color space) between the network's predicted color of the reflectance (the output) and the "true color" (the color under neutral illumination with a neutral background) is less than the difference between the "true color" and the "physical color" (the color as calculated from the power spectrum of the reflectance times the illuminant). "Physical color" is what would be expected in a completely non-color constant system, such as a camera or photometer. A *"shift toward constancy"* is a shift of the output toward the true color and away from the physical color.

When evaluating the behavior of the network the "true color" response is defined as the output of the network for a reflectance under standard neutral illuminant and background conditions. We call the "true color" of a reflectance an "eigencolor" of the system because it is an internal reference which is not altered or "corrected" by the system in the way that those inputs which have non-neutral backgrounds or illuminants are. The "eigencolor" response represents "perfect constancy". A reference was also needed for a non-constant response, which would correspond to the "physical color" reference, reflectance times illuminant.

In human perception, stimuli which are viewed "in isolation", either with a completely black background or viewed at a distance through a hole or aperture in a grey barrier, result in a close correlation between the physical color and the perceived color. In other words, in "aperture viewing mode" color constancy disappears. ([29], also called "void viewing mode" in [59], similar to "film colors" reviewed in [1]) Without a surround which also reflects the illuminant change, the visual system cannot determine whether a change in the color signal from the test spot is due to an illuminant change or a reflectance change.

Stimuli analogous to the aperture viewing condition were created for the network by simulating a change in the illuminant on only the center test spot, leaving the background as a neutral reflectance under a neutral illuminant. (See figure 4) As with a human viewing a color in aperture mode, the outputs to these stimuli could be considered essentially *non*-color-constant, because there is no change in the surround color signal to indicate that the change is due to a different illuminant and not to a change in reflectance.

The "aperture colors" will not be perfectly non-constant because of a small amount of compression from the adaptation of those receptors directly on the test spot. Therefore the method underestimates slightly the contribution of adaptation in the simulation, but the effect is minor compared to the size of the color constancy shifts under natural viewing conditions. Simulation stimuli for "natural viewing conditions", in which the illuminant was applied to the entire image, are the test stimuli.

The Euclidean distance in output space $\Delta(O_R{}^2 + O_G{}^2 + O_B{}^2)^{1/2}$ was calculated between each aperture color response and the corresponding eigencolor response. This distance was then compared to the distance between the test stimulus output and the eigencolor output. The percent constancy achieved by the network is calculated by:

$$\left(\frac{D_{ea} - D_{et}}{D_{ea}}\right) \times 100 \qquad (8)$$

where D_{ea} is the distance from eigencolor to aperture color response, and D_{et} is the distance from eigencolor to test stimulus response. In this way, a quantitative measure was obtained for the network's ability to discern the "true" color of the center reflectance by using information about the illuminant from the surround. We will use this measure to assess and compare the effectiveness of the network in different simulated stimulus and background conditions. Obviously, the end points with 0% and100% correspond exactly to their psychophysical counterparts. Correspondence of computed values between 0% and100% cannot be taken as numerically equivalent to psychophysics. Psychophysical color space is not Euclidean nor linear in R G B along the curve from 0% to 100% and the computed distance measures may not exactly match.

ii. Color Constancy with a Homogeneous Neutral Background

(a) Natural Illuminants

Because of the importance for practical applications, and for evolutionary significance, we begin by looking at the color constancy abilities of the network under natural illuminants, such as the various phases of daylight. The power spectra used are from tabulations of the CIE standard illuminants, A, B, C, D55, D65, and D75 [56]. We will refer to them here by qualitative descriptions of their sources: incandescent light (or full radiator), direct sunlight, average daylight (averaged over all times of day from dawn to dusk), overcast daylight, natural daylight (a single phase of daylight), blue skylight. These spectra are shown in figure 5. The natural illuminants vary primarily in the short (blue) and long (red) wavelength regions with very little change in the middle (green) wavelengths.

The network outputs for three Munsell reflectances (2.5yr7/10, 10b6/10, and 2.5bg6/10,) under these illuminants are shown in figure 6. The stimuli consisted of a single small square reflectance patch (the "test spot") in the center of an homogeneous grey background reflectance. The figure shows the ratio of the network outputs in a ternary plot. The plots shown are sub-spaces of the full graph in which the ratios between each of the outputs and the sum of all three outputs range from 0 to 1. Each of the three corners of the full triangular graph represents activity in a single channel only, as would result from a monochromatic light stimulating one color channel exclusively. Because we are using broadband reflectances and illuminants and because the cone spectral sensitivity functions overlap significantly, the network outputs are concentrated in the center of the graph. In particular, the $O_R/(O_R+O_G+O_B)$ and $O_G/(O_R+O_G+O_B)$ ratios are usually close to 0.5 because of the large amount of overlap in the R and G cone sensitivities.

The squares mark the eigencolor outputs. The outputs to the aperture colors are shown with solid circles. As explained in the previous section, these are almost completely non-constant, and represent the physical change in the color signal. The open circles represent the test stimuli outputs, the responses to each of the reflectances under each of the illuminants under natural viewing conditions. The network responses to the test stimuli are significantly closer to the eigencolor outputs than the aperture color outputs are to the eigencolor outputs, demonstrating color constancy.

(b) Larger Illuminant Changes

Because natural illuminants are relatively similar to each other, the constancy shifts required in the previous sections were rather small. To see how the network would handle larger shifts, more strongly colored illuminants were tried. These illuminants were a linear combination of a spectrally flat illuminant and one of three illuminants with Gaussian spectral distributions which peaked at 440nm, 560nm, and 660nm. Some examples are shown in figure 7. Solid black circles represent aperture color outputs, and open circles represent test stimuli outputs. Eigencolor outputs are shown by squares.

To see the difference in the system's color constancy abilities in natural versus artificial illuminants, compare figure 6 to figure 7. Although the percent constancy achieved with natural and artificial illuminants is similar (average shift for all stimuli was approximately 50%), the total distances, caused by the physical color change, for which compensation is needed, are much smaller for the natural illuminants than for the artificial illuminants. Therefore, the variance in the test stimuli outputs is much smaller for the natural illuminant conditions. Because the biological system evolved to deal with natural

illuminants, it should not be surprising that the amount of constancy achieved in laboratory conditions with artificial illuminants and viewing conditions is relatively small.

iii. Color Constancy with Complex Backgrounds

The network implementation of this color constancy algorithm allows complex backgrounds to be handled as easily as simple backgrounds. This is not to say that the response to a complex background is the same as the response to a spatially weighted average of the background (the equivalent surround hypothesis: [49]). As has been demonstrated in psychophysical experiments [7, 27], local features in the background region of the image, such as low contrast edges, can affect the constancy and induction results for the test spot. In the simulation, color and luminance edges in the background influence responses of the local operations in the retinal layers of the network. For example, the spectrally opponent cells enhance luminance edge responses. These enhanced responses are then spatially integrated and used by the spectrally specific contrast operation in the final stage of the network.

The network was tested using Mondrian images, in addition to the grey homogeneous background condition described in the previous section. Some examples are shown in figure 8. Again, solid circles represent aperture color outputs and open circles represent the test stimuli outputs. Sqaures represent the outputs for those reflectances with the Mondrian background and a neutral illuminant. Therefore, the squares represent the color constant output, but do not eliminate color induction effects from the Mondrian background reflectances.

The triangles mark the eigencolor outputs for these two reflectances. These outputs represent each of the reflectances under neutral illumination with a homogeneous neutral background. The shifts from the neutral background, eigencolor outputs to the square symbols, show the color induction due to the Mondrian background. The colors of the patches in the Mondrian used for the results shown here were chosen at random and had a yellow average chromaticity. Therefore, independent of the influence of the illuminants, there is a shift in the outputs away from red and green and toward blue. The amount of induction depends on the amount of color contrast between the center test spot and the background. Color induction will be discussed in more depth in section *v*.

iv. Gradients in Illumination

Most previous computational algorithms for color constancy required that the illuminant be constant across space (e.g., [5], [8], [14]) or at least that the illuminant vary slowly relative to the reflectance changes [15, 31, 42]. This assumption was necessary in

order to make the computations tractable and to distinguish changes in illuminant from changes in reflectance. However, nearly all natural scenes have gradients in the illumination from multiple light sources, varying distance to the light source, and shadows. The neural network's distributed representation of the input image does not have this requirement. We tested the network using simulated images (test spot plus uniform grey background) with chromatic gradients in the illuminant.

There is little psychophysical data on the effects of gradients, either in luminance or color, on the perception of color. Therefore, we tested our network on simple stimuli which would have an easily predictable perceptual effect. We chose to use circularly symmetric images with symmetric illuminant gradients. The linear gradients were either horizontal or vertical across the image and were pivoted about the center of the image so that the gradient was positive on one side of the central test spot and negative on the other side. Such a stimulus would be expected to have no net effect on the perception of the color at the center of the image.

The spatially homogeneous illuminants used above are simulations of a illuminants passed through spatially homogeneous filters. The illuminant gradients were effectively a simulation of these same illuminants, but passed through one or more filters whose density increases linearly across the image. The gradient filter has the same average density as the spatially homogeneous filter used above. The activation levels of cones of one type were multiplied by $(1 + (a * d))$, where a is the "amplitude" of the gradient and d is the distance (either positive or negative) from the center of the image along either the horizontal or vertical axis. The distance was measured as the distance from the center of the image to the center of the cone's receptive field as a fraction of the distance from the center of the image to the edge of the image. A constant value was added or subtracted uniformly across the entire image for some of the stimuli to accomodate a DC offset.

The results are shown in figure 9. Each group of data points represents a different offset value. The horizontal axis shows the color channel that contained the gradient, and the amplitude of the gradient multiplication factor at it's highest point (the outer edge of the image). As the figure shows, the offsets produced a shift in the R/G ratio of the outputs, but the gradients showed no effect at all on the output of the network, either for chromatic or luminance gradients. This result would be expected intuitively for human perception as well because the background was uniform and the gradients were symmetric. The increase in activity on one side of the image is canceled by the decrease in activity on the other side of the image when the large surrounds of the V4 network stage integrate their inputs. Had this image contained a more complex background, which was not spatially symmetric in the chromaticities of the reflectances, then a chromatic or

luminance bias could have been introduced by the gradient. The important point here is that the network enables the computation of color in the presence of gradients.

v. Color Induction

In their implementation of the retinex algorithm, Moore *et al.* [39] were concerned about the existence of color induction in their output images. Indeed, if the induced shift in color output is too great, or in the wrong color direction, then color induction is undesirable. However, if we wish to imitate human color vision, our system must have color inducing behavior. In addition, color induction is helpful if we wish to take advantage of the increase in contrast that color induction creates in order to do object detection or surface segmentation.

The current network simulation produces color induction through the same mechanisms which produce color constancy. Color induction results are shown in figure 10. The size of the induction shift increases as the size of the surround increases and decreases as the width of the gap between the inducing surround and the center test spot increases. The size of the induction shift also depends on the color difference between center and surround as can be seen in the different induction amplitudes for the bluegreen center, blue surround stimulus and the purple center, red surround stimulus. This difference in amplitude is because of the dependence of the final stage of the network on the spectrally specific contrast between center and surround. If the contrast between center and surround is small, then the push-pull spectrally-specific contrast mechanism will not be activated. On the other hand, if the difference between center and surround is large, then one or more of the color channels for either the center, the surround, or both, is likely to be responding in the nonlinear region of its response curve where the gain is small, making the induction ineffective.

vi. Color Assimilation

Another practical consideration for color applications is color assimilation, which is the opposite of color induction. Assimilation results in the blending of colors, a decrease rather than an increase in color contrast. Assimilation is usually seen when a fine pattern is placed on a colored background. For example, thin white stripes over a colored background will make the background appear lighter. Thick white stripes, however, will cause induction and make the background appear darker.

Color assimilation occurs in the simulation as a result of the spectrally and spatially opponent cells in the second layer of the network. As explained in section *iv* of "Network Architecture", the opponent stage enhances luminance edges, but diminishes

chromatic edges. This is caused by the difference in spectral sensitivities between the center and surround of these cells' receptive fields. All three cone types have responses which are highly correlated to a change in luminance, so the spectrally opponent sensitivity of the cell is not a factor at a luminance edge. A high spatial frequency change in color, however, often causes decorrelated responses between cone types, leading to the decrease in cell responses near a chromatic border when the excitatory center receives different input than the inhibitory surround. For example an R+G- cell whose center is stimulated by a red region in an image will have a smaller response if its receptive field is close enough to the edge of a green region so that the green inhibitory surround is stimulated by the green region in the input image.

The response of the R on-center opponent layer to an assimilation stimulus is shown in figure 11. In this example, a red spot and an open red square are shown with a yellow background. When the spot is large and the lines of the square are wide so that both the excitatory center and the inhibitory surround portions of the opponent cell receptive fields are stimulated by the same color region, the color gain of the network is high. However, when the spot is small and the lines of the square are narrow so that the center and the surround regions of the opponent cells each receive different color inputs, the color gain of the network is low. The difference between the network response to the pattern and the response to the background is greater for the low spatial frequency stimulus than for the high spatial frequency stimulus. This is consistent with human color perception. As with color induction, color assimilation must be considered if one wishes to make a system which predicts human color perception.

DISCUSSION

We have shown that by using mechanisms similar to those found in the human visual system, one can create an artificial system which has good color constancy, flexibility in the types of images that it can handle, and very few assumptions about the images. Both simple and relatively complex backgrounds can be used and the color constancy behavior of the network simulation is not affected by gradients in the illuminant. In addition, because the network calculates its own reference level, it does not require the "grey world assumption", which is that the average reflectance chromaticity is the same for all images. As is true with the human visual system, the network does not produce perfect color constancy but only a shift towards constancy. Exact imitation of the human visual system, however, will require consideration of additional complex features, as will be discussed below, and necessitate more complex neural circuitry and processing stages not available in the present network.

Color induction is produced by the network in addition to color constancy. Although this could be interpreted as undesirable for some systems because it affects the accuracy of the color output, in many applications this could be an advantage because it enhances color contrast and therefore aids in object detection. In addition, because the magnitudes and directions of the color shifts are consistent with human color psychophysical data, the color induction behavior of the network would be needed in a system intended to mimic human color perception. Another aspect of human color perception, color assimilation, is also produced by the simulation in the spectrally opponent layer of the network. The human visual system also displays luminance assimilation, although the magnitudes and spatial scales of color and luminance assimilation are somewhat different [1]. Luminance assimilation is not produced in the current network because of the enhancement, rather than diminishment, of high spatial frequency luminance edges. In the biological system, luminance assimilation may be caused by factors not included in the current simulation such as light scatter and excitatory lateral neural connections such as cone-cone gap junctions.

This model has several benefits in overcoming restrictions associated with the computation of color constancy. The concept of spectrally-specific contrast, included in the current model, is similar to the independent "lightness" channels of the Retinex algorithm [31, 36, 38, 48]. Several "lightness" algorithms which have been proposed for color constancy, have been shown to be mathematically equivalent to a local spatial derivative plus a normalization term [23]. Some have argued (see review [26]) that the Retinex is similar to Von Kries adaptation in that each is a renormalization of color channel activities relative to some white reference. In that sense, the adaptation stage and the spectrally specific contrast stage in this simulation are also similar. However, there are several important differences which allow the operations described here to each cover different stimulus conditions and to cooperate in their contributions to color constancy.

First, the cone adaptation stage has a permanent white reference set by the midpoint of the range of possible threshold values. The adaptation stage also has a long-term adaptation reference which is usually close to neutral because it is established through exposure to many different stimuli over a long period of time. Faster, more localized adaptation effects are deviations from this long term reference. The reference for the spectrally specific contrast stage is the activity of the "local reference" cells. The spectrally specific contrast reference is not fixed, but instead changes with each new image. This reference is also not usually neutral. The cone specific contrast mechanism described here is also different from most "lightness" algorithms in that the normalizing

reference is measured locally, rather than a globally. The local reference is used in a push-pull mechanism with a spatially weighted global measure of spectrally specific contrast.

Second, the spatial profiles of the adaptation and V4 mechanisms are different. The effect of localized adaptation depends more heavily on the central test spot, while the effect of the large surrounds in the spectrally specific contrast operation are more affected by the background stimulus. The effect of the spectrally specific contrast operation is increased by the spectral and spatial opponency of the preceding stage which enhances color and brightness contrast.

There are several improvements to be made to the network. An obvious next step is to use "real world" images. There are two additions that need to be made to the network in order to allow it to handle real images appropriately. The first is spatial scale invariance. In the human visual system, the color of an object does not depend on the size of that object, provided that the surrounding objects are also scaled proportionately. In the current network simulation, however, the size of the test color patch must be appropriate for the size of the receptive fields of the cells in each of the network layers. One possible solution would be to incorporate dynamic receptive fields into the network which would change to match the size of each uniform color region in the image.

Another possible solution to the scale invariance problem would also incorporate the second aspect of real world imagry which remains to be addressed, that of surface segmentation and depth perception. Color and brightness perception are known to depend on the perceived depth plane of the object of interest relative to its surrounding surfaces, with color constancy and induction effective only within a single depth plane [45, 55]. Feedback from a surface segmentation process which assigns a relative depth to each surface, such as that described by Finkel and Sajda [16] could enhance or inhibit the color inducing effects of each region within the final stage of the network.

Each of the stages in the network contributes to the overall color constancy behavior. The final stage is a novel mechanism based on the chromatic properties of cells found in cortical area V4. The explicit calculation of a local reference and the use of a push-pull mechanism to incorporate the contrast responses in the final output, in addition to producing color constant responses for most images, allows the network to give accurate color responses for images which contain little chromatic contrast. The adaptation stage increases the dynamic range of the system, increases its flexibility to incorporate different viewing conditions, and also enhances the color constancy abilities of the system. The middle stage, the spectrally and spatially opponent mechanism, serves to enhance contrast, and therefore constancy, particularly at high spatial frequency luminance

edges. Together, the stages of the model comprise a system which produces good color constancy while maintaining flexibility.

Acknowledgements: This work was supported by ONR grants N00014-93-1-0681, by AFOSR grants NL 91-0082 and 92-J-0316, and by grants from The Whitaker Foundation, and The McDonnell-Pew Program in Cognitive Neuroscience.

References

[1] Beck, J. *Surface Color Perception,* Cornell University Press: London, 1972.

[2] Brainard, D.H. and Wandell, B. "Analysis of the retinex theory of color vision," *Journal of the Optical Society of America A* , vol.3, pp. 1651-1661, 1986.

[3] Brainard, D.H. and Wandell, B. "A bilinear model of the illuminant's effect on color appearance," in *Computational Models of Visual Processing,* J.A. Movshon and M.S. Landy eds, MIT Press: Cambridge, MA, 1991.

[4] Brainard, D.H. and Wandell, B. (1992) "Asymmetric color matching: How color appearance depends on the illuminant," *Journal of the Optical Society A* 9, 9: 1433-1448.

[5] Brill, M.H. "A device performing illuminant-invariant assessment of chromatic relations," *Journal of Theoretical Biology,* vol. 71, pp. 473-478, 1978.

[6] Brill, M.H. and West, G. "Chromatic adaptation and color constancy: A possible dichotomy," *COLOR research and application* , vol. 11, pp. 196-204, 1986.

[7] Brown, R.O. "Integration of enhanced-contrast edges in color vision," *Investigative Ophthalmology & Visual Science*, vol. 34, p. 766, 1993.

[8] Buchsbaum, G. "A spatial processor model for object colour perception," *Journal of the Franklin Institute* , vol. 310, pp. 1-26, 1980.

[9] Cornellissen, F.W., and Brenner, E. "On the role and nature of adaptation in chromatic induction," in *Channels in the Visual Nervous System: Neurophysiology, Psychophysics and Models* (ed. B. Blum) p. 109-124, 1991.

[10] Desimone, R. and Schein, S.J. "Visual properties of neurons in area V4 of the macaque: Sensitivity to stimulus form," *Journal of Neurophysiology* , vol. 57, pp. 835-868, 1987.

[11] Dannemiller, J.L. "Rank orderings of photoreceptor photon catches from natural objects are nearly illuminant-invariant," *Vision Research,* vol. 33, pp. 131-140, 1993.

[12] Dufort, P.A., and Lumsden, C.J. "Color categorization and color constancy in a neural network model of V4," *Biological Cybernetics,* vol. 65, pp. 293-303, 1991.

[13] D'Zmura, M. and Iverson, G. "Color structure from chromatic motion. II. Results for two-stage linear recovery," *Journal of the Optical Society of America,* in press.

[14] D'Zmura, M. and Lennie, P. "Mechanisms of color constancy," *Journal of the Optical Society of America A* , vol. 3, no. 10, pp. 1662-1672, 1986.

[15] Faugeras, O.D. "Digital color image processing within the framework of a human visual model," *IEEE Transactions on Acoustics, Speech and Signal Processing* , vol. 27, pp. 380-393, 1979.

[16] Finkel, L. H. and Sajda, P. "Object discrimination based on depth-from-occlusion," *Neural Computation,* vol. 4, pp. 901-921, 1992.

[17] Gershon, R. and Jepson, A.D. "The computation of color constant descriptors in chromatic images," *COLOR research and application* , vol. 14, pp. 325-334, 1989.

[18] Grossberg, S. "Cortical dynamics of three-dimensional form, color and brightness perception," *Perception and Psychophysics,* vol. 41, no. 2, pp. 87-158, 1987.

[19] Hayhoe, M.M, Benimoff, N.I. and Hood, D.D. (1987) "The time-course of multiplicative and subtractive adaptation process," *Vision Research* 27,1981.

[20] Hayhoe, M. and Wenderoth, P. (1991) "Adaptation mechanisms in color and brightness," in *From Pigments to Perception,* (eds. A. Valberg and B. Lee) Plenum Press, 353-367.

[21] Helmholtz, H. von *Physiological Optics* (Trans. J.P.C. Southall, 1924) vol. 2, pp. 286-87, Rochester, NY: Optical Society of America, 3rd edition, 1866.

[22] Hubel, D.H. and Wiesel, T.N. (1968) "Receptive fields and functional architecture of monkey striate cortex," *Journal of Physiology (London)* 195, 215-243.

[23] Hurlbert, A. "Formal connections between lightness algorithms," *Journal of the Optical Society A* , vol. 3, no. 10, pp. 1684-1693, 1986.

[24] Hurlbert, A.C. and Poggio, T.A. "Synthesizing a color algorithm from examples," *Science,* vol. 239, pp. 482-485, 1988.

[25] Ingle, D.J. "The goldfish as a retinex animal," *Science,* vol. 227, pp. 651-654, 1985.

[26] Jameson, D. and Hurvich, L.M. "Essay concerning color constancy," *Annual Review of Psychology,* vol. 40, pp. 1-22, 1989.

[27] Jenness, J.W., and Shevell, S.K. "Chromatic complexity and color appearance: A dark cloud over the gray world hypothesis," *Investigative Ophthalmology and Visual Science,* vol. 34, pp. 765, 1993.

[28] Judd, D.B. (1951) "Report of U.S. Secretariat Committee on Colorimetry and Artificial Daylight," *CIE Proceedings,* Vol. 1, Part 7, p. 11, Paris: Bureau Central de la CIE.

[29] Judd, D.B. "Appraisal of Land's Work on Two-Primary Color Projections," *Journal of the Optical Society of America,* vol. 50, pp. 254-268, 1960.

[30] Katz, D. and Revesz, G. "Experimentelle studien zur vergleichenden psychologie (Versuch mit hulmetn)," *Zeits f. angew. Psychol.,* vol. 18, p. 307, 1921.

[31] Land, E.H. "The retinex theory of colour vision," *Proceedings of the Royal Institute of Great Britain,* vol. 47, pp. 23-58, 1974.

[32] Land, E.H. " Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image," *Proceedings of the National Academy of Science,* vol. 80, pp. 5163-5169, 1983.

[33] Land, E.H. "Recent advances in Retinex theory," *Vision Research,* vol. 26, pp. 7-21, 1986.

[34] Land, E.H. and McCann, J.J. "Lightness and retinex theory," *Journal of the Optical Society of America,* vol. 61, no. 1, pp. 1-11, 1971.

[35] Lennie, P., and D'Zmura M. "Mechanisms of color vision," *Critical Reviews in Neurobiology* , vol. 3, no. 4, pp. 333-401, 1988.

[36] Lucassen, M. and Walraven, J. "Quantifying color constancy: Evidence for nonlinear processing of cone-specific contrast," *Vision Research,* vol. 33, no. 5/6, pp. 739-757, 1993.

[37] Maloney, L.T. and Wandell, B.A. "Color constancy: A method for recovering surface spectral reflectance," *Journal of the Optical Society of America,* vol. 3, pp. 29-33, 1986.

[38] McCann, J.J., McKee, S.P., and Taylor, T.H. "Quantitative studies in Retinex theory: A comparison between theoretical predictions and observer responses to the 'Color Mondrian' experiments," *Vision Research* , vol. 16, pp. 445-458, 1976.

[39] Moore, A., Allman, J., and Goodman, R.M. "A real-time neural system for color constancy," *IEEE Transactions on Neural Networks,* vol. 2, no. 2, pp. 237-246, 1991.

[40] Moran, J., Desimone, R., Schein, S.J., and Mishkin, M. "Suppression from ipsilateral visual field in area V4 of the macaque," *Society for Neuroscience Abstracts,* vol. 9, pp. 957, 1983.

[41] Naka, K. and Rushton, W.A.H. "S-potentials from luminosity units in the retina of fish (Cyprinidiae)," *Journal of Physiology,* vol. 188, pp. 587-599, 1966.

[42] Rubin, J.M. and Richards, W.A. "Color vision and image intensities: When are changes material?" *Biological Cybernetics,* vol. 45, pp. 215-226, 1982.

[43] Sajda, P. and Finkel, L.H. "NEXUS: A simulation environment for large-scale neural systems," *Simulation,* vol. 59, no. 6, pp. 358-364, 1992.

[44] Schein, S.J. and Desimone, R. "Spectral properties of V4 neurons in the Macaque," *Journal of Neuroscience,* vol. 10, no. 10, pp. 3370-3389, 1990.

[45] Schirillo, J.A. and Shevell S.K. "Perceived brightness, but not lightness is influenced by retinally non-adjacent coplanar surfaces," *OSA Annual Meeting Technical Digest* , vol. 23, p. 51, 1992.

[46] Schnapf, J.L., Nunn, B.J., Meister, M., Baylor, D.A. "Visual transduction in cones of the monkey macaca fascicularis," *Journal of Physiology,* vol. 427, pp. 681-713, 1990.

[47] Tiplitz Blackwell, K. and Buchsbaum, G. "The effect of spatial and chromatic parameters on chromatic induction," *COLOR research and application* vol. 13, no. 3, pp. 166-173, 1988.

[48] Tiplitz Blackwell, K. and Buchsbaum, G. "Quantitative studies of color constancy," *Journal of the Optical Society of America,* vol. 5, pp. 1772-1780, 1988.

[49] Valberg, A. and Lange-Malecki, B. "'Color constancy' in Mondrian patterns: a partial cancellation of physical chromaticity shifts by simultaneous contrast," *Vision Research,* vol. 30, pp. 371-380, 1990.

[50] Von Kries, J. "Die Gesichtsempfindungen," in W. Nagel (ed.) *Handbuch der Physiologie der Menschen*, pp. 109-282. Vieweg, Brunswick, 1905.

[51] Vos, J.J. "Colorimetric and photometric properties of a 2° fundamental observer," *COLOR Research and Application,* vol. 3, p. 125-128, 1978.

[52] Vos, J.J., and Walraven, P.L. "On the derivation of the foveal receptor primaries," *Vision Research,* vol. 11, pp. 799-818, 1971.

[53] Walraven, J., Enroth-Cugell, C., Hood, D.C., MacLeod, D.I.A., and Schnapf, J.L. "The control of visual sensitivity: Receptoral and postreceptoral processes," in *Visual Perception: The Neurophysiological Foundations*, (ed. L. Spillman and J.S. Werner) Academic Press, Inc: San Diego, CA, 1990.

[54] Wesner, M.F. and Shevell, S.K. "Color perception within a chromatic context: changes in red/green equilibria caused by noncontiguous light," *Vision Research,* vol. 32, no. 9, pp. 1623-1634, 1992.

[55] White "A new effect of pattern on perceived lightness," *Perception,* vol. 8, pp. 413-416, 1979.

[56] Wyszecki, G. and Stiles, W.S. *Color science* (2nd edn.) Wiley: New York, 1982.

[57] Zaidi, Q., Yoshimi, B., Flanigan, N. and Canova, A. "Lateral interactions within color mechanisms in simultaneous induced contrast," *Vision Research,* vol. 32, no. 9, pp. 1695-1708, 1992.

[58] Zeki, S.M. "Colour coding in the cerebral cortex: The reaction of cells in monkey visual cortex to wavelengths and colors," *Neuroscience,* vol. 9, pp. 741-765, 1983.

[59] Zeki, S.M. "Colour vision and functional specialisation in the visual cortex," *Discussions in Neuroscience,* vol. 6, no. 2, pp. 11-64, 1990.

[60] Zrenner, E. *Neurophysiological aspects of color vision in primates: Comparative studies on simian retinal ganglion cells and the human visual system.* New York: Springer, 1983.

Figure Captions

FIG 1 Detail of the cortical (top) and retinal (bottom) stages of the simulation. White triangles represent on-center cells, shaded triangles are off-center cells, and large circles are interneurons. Synapses are shown by small ovals, in white for excitatory, black for inhibitory. The silent surrounds in the cortical layers have an exponential synaptic weighting function as is shown inside the interneuron providing input to each contrast cell. In the retinal layers, the adaptation mechanism receives Gaussian weighted input from nearby cones. The adaptation cell then provides feedback to the cone at the center of its receptive field, changing that cone's threshold. On- and off-center cells in the retina have a difference of Gaussians synaptic weighting profile, as is shown inside each cell.

FIG 2 The Vos-Walraven cone spectral sensitivity functions. Figure adapted from [52].

FIG 3 Response curves for cones in the simulation under a range of values for the adaptation threshold. **(a)** Shows the sigmoidal limits of the adaptation range. The luminance level of the adapting stimulus was increased linearly, but the threshold values reach an asymptote at both ends of the range. **(b)** Same as (a) in log-linear coordinates.

FIG 4 Illustration of the simulated method for producing "aperture colors". The aperture color illuminant condition has the standard neutral illuminant on every part of the image except the test spot. The test spot is illuminated exclusively by the test illuminant, "illuminant 2". In the natural viewing condition, there is a single source which illuminates the entire image.

FIG 5 The power spectra of several natural illuminants [56]. The solid line is a full radiator of color temperature 2856K, similar to incandescent lighting. The others are all phases of daylight. Long dashes represent direct sunlight; short dashes, natural daylight averaged over all times of day; dotted line, overcast skylight; long odd dashed, natural daylight; and short odd dashed, clear blue sky. The illuminants are all very similar in the middle wavelengths but vary in their relative amounts of power in the long and short wavelengths.

FIG 6 Color constancy shifts made by the network for three Munsell reflectances (2.5yr7/10, 10b6/10, 2.5bg6/10) with the illuminants shown in figure 6. The axes show the relative proportions of each of the three network outputs. The plots shown are sub-

spaces of the full graph in which the ratios between each of the outputs and the sum of the other two range from 0 to 1. Each of the three corners of the full triangular graph represents activity in a single channel only, as would result from a monochromatic light stimulating one color channel exclusively. The squares represent the eigencolors of the reflectances, the response of the network with neutral illumination. The solid circles are the non-constant responses to aperture colors as explained in "Results: i. measuring constancy". The displacements of these symbols, from the square in each graph, represent the physical change in the color signal as the illuminant is changed. The aperture color closest to the O_B corner of the graph is for the blue skylight illuminant, and the aperture color closest to the O_R corner of the graph is for the incandescent illuminant. Responses for all other illuminants fall between these two extremes. The open symbols show the test stimulus responses of the network in natural viewing conditions. The open symbols show a shift toward constancy, away from the aperture colors and toward the eigencolors. Arrows show the direction of the shift.

FIG 7 Color constancy shifts for Munsell reflectances 2.5bg6/10, 2.5g7/6, 2.5yr7/10, and 5.0r5/12. The illuminants used here are the combination of a spectrally flat illuminant and one of three illuminants with Gaussian spectral distributions peaking at 440nm, 560nm, and 660nm. The physical color shifts produced by these illuminants (distance from eigencolor to aperture color) are larger than those produced by natural illuminants in figure 7. Therefore, although the percent color constancy shifts are similar for the natural and artificial illuminants, the variances for the test stimuli outputs are smaller for the natural illuminants. As in figure 7, solid symbols mark aperture colors and open symbols mark test stimuli responses. Eigencolors are shown by the squares. Arrows are shown to indicate the direction of shift the reflectance-illuminant pairs.

FIG 8 Color constancy with a complex background, a Mondrian, for two reflectances, 10gy7/10 and 10b6/10 which were placed at the center of the Mondrian. Responses are shown for incandescent light (A, shown in fig. 5) for both reflectances, for direct sunlight (B) with reflectance 10gy7/10 and for blue skylight (D75) with reflectance 10b6/10. Solid and open symbols are as in figs. 6 and 7. The squares are the neutral illuminant, Mondrian background outputs and so represent the color constant response, but do not eliminate the color induction effects of the Mondrian background. The triangles mark the eigencolors, the responses of the network under neutral illumination with a neutral homogeneous background. The differences between the neutral illumination, Mondrian backgound outputs and the eigencolor responses show the induction effect of the

Mondrian background, independent of illumination changes. Because the test responses were obtained with the Mondrian background, the constancy shifts are toward the neutral illuminant, Mondrian background responses.

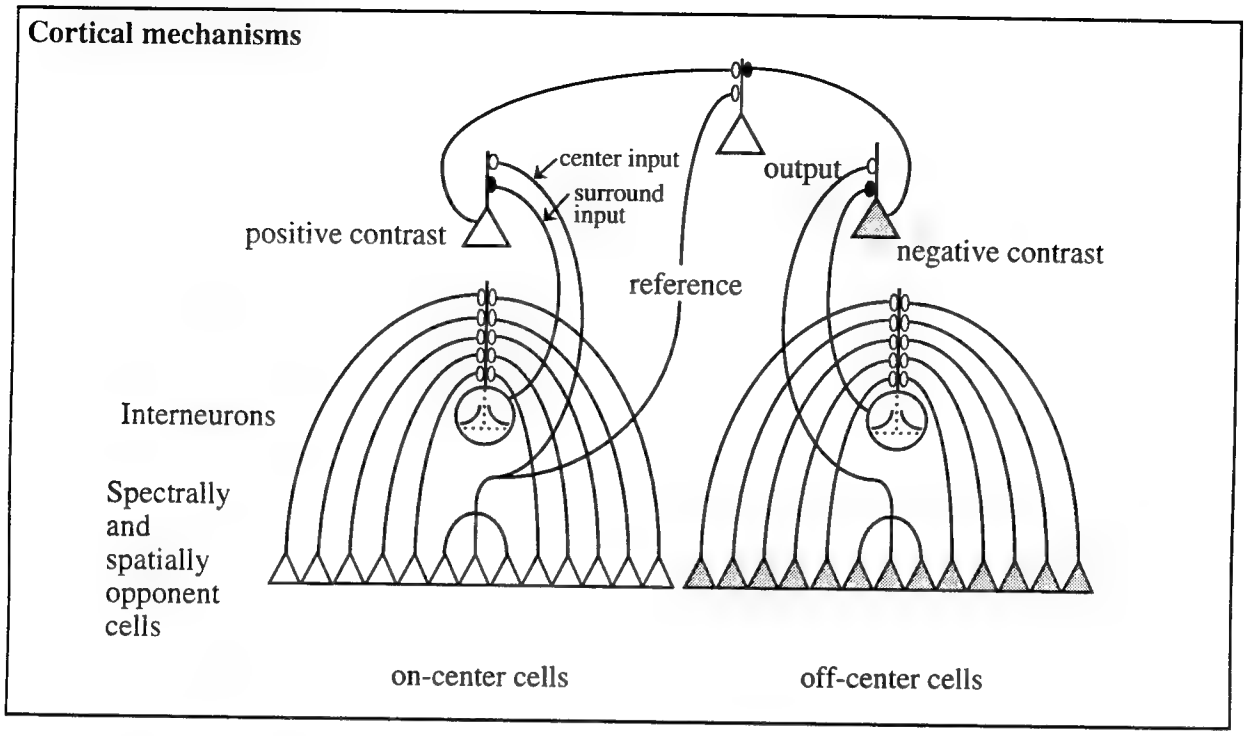
FIG 9 Change in output ratios of the network with various offsets and gradient amplitudes in the input image. The results show a shift toward R_{out} for a positive R offset and a shift toward G_{out} for a positive G offset, as expected. Because all operations in the network are circularly symmetric, there is no effect from the addition of a symmetric linear illuminant gradient to the stimulus.

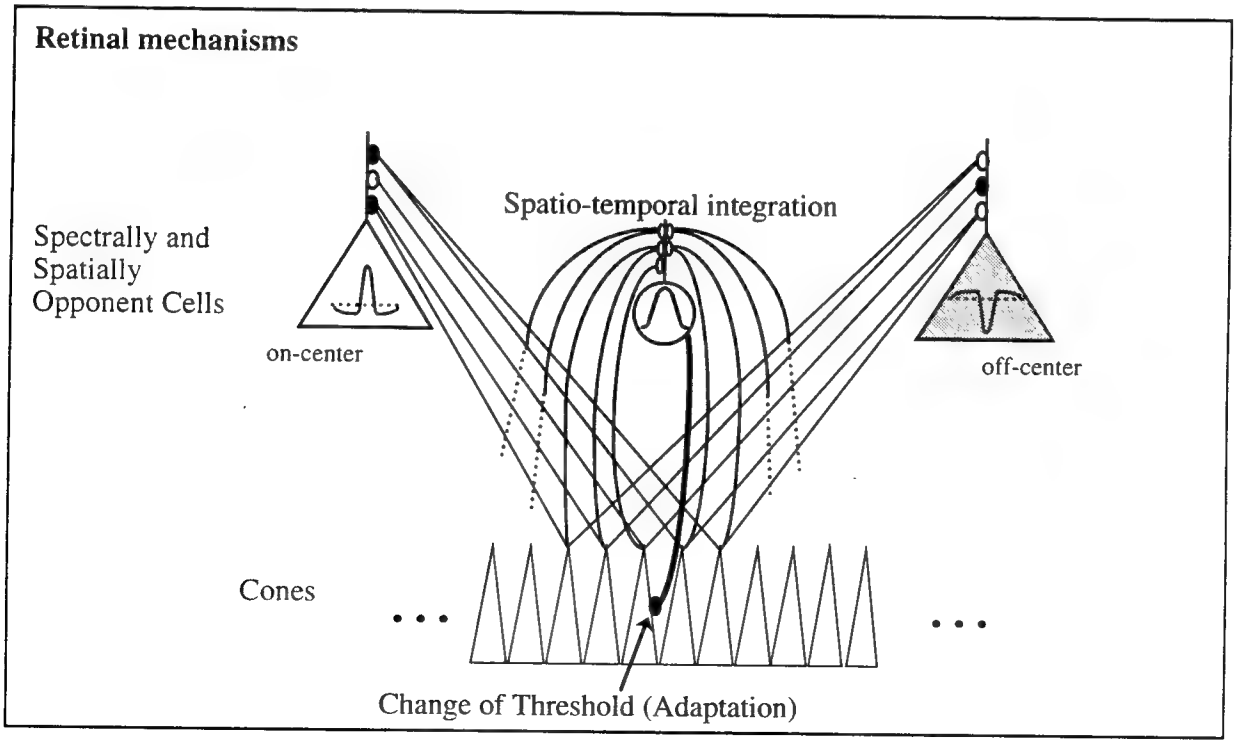
FIG 10 **(a)** Simulation image used to measure the spatial properties of color induction. Width of the gap and of the surround can be varied. The center patch is 3x3 input units, the same size as the centers of the V4 receptive fields. **(b)** Results for two examples of color induction. The y-axis shows the size of the induction shift measured as a change in the ratio of two of the outputs due to the presence of the inducing surround. For the circles, the center reflectance is blue-green and the surround is blue. For the diamonds, the center is purple and the surround is red. As the gap between center and surround is increased, the induction effect decreases. At 4 units separation, the surround of the stimulus is outside of the silent surrounds of the V4 cells. **(c)** With no gap, the width of the surround is varied. As the width of the surround increases, the amount of induction increases. The amplitudes of the induction shifts for the two images are dependent upon the contrast between the center and surround colors.

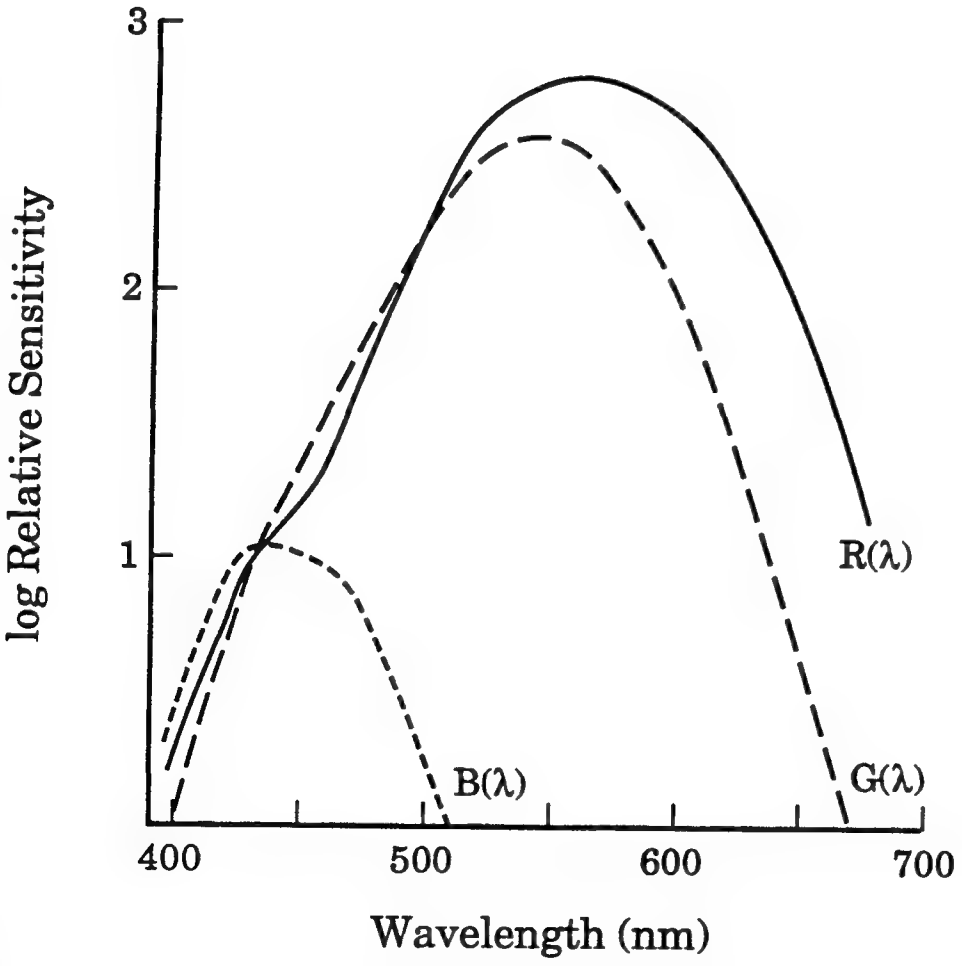
FIG 11 Demonstration of color assimilation in the spectrally opponent layer of the network. The input stimulus for each plot is a red spot and a red square with a yellow background. The two stimuli vary only in the width of the lines comprising the square and the size of the central spot. On the right, the coarse pattern (the wider lines and larger spot) results in a higher contrast in the response of the network layer cells. With the fine pattern (left), the contrast is reduced. This can also be seen in the cross-sectional profiles of the response of the network layer. The fine pattern is shown by the dashed lines and the coarse pattern by the solid lines. The background near the square has a higher R response (looks more red) and the pattern itself has a lower R response (looks more like the yellow background).

Table 1: *Each of the most significant parameters in the simulation is presented along with the criteria used to determine that parameter's value. In parentheses are the values used for the results presented here and the range of possible values.*

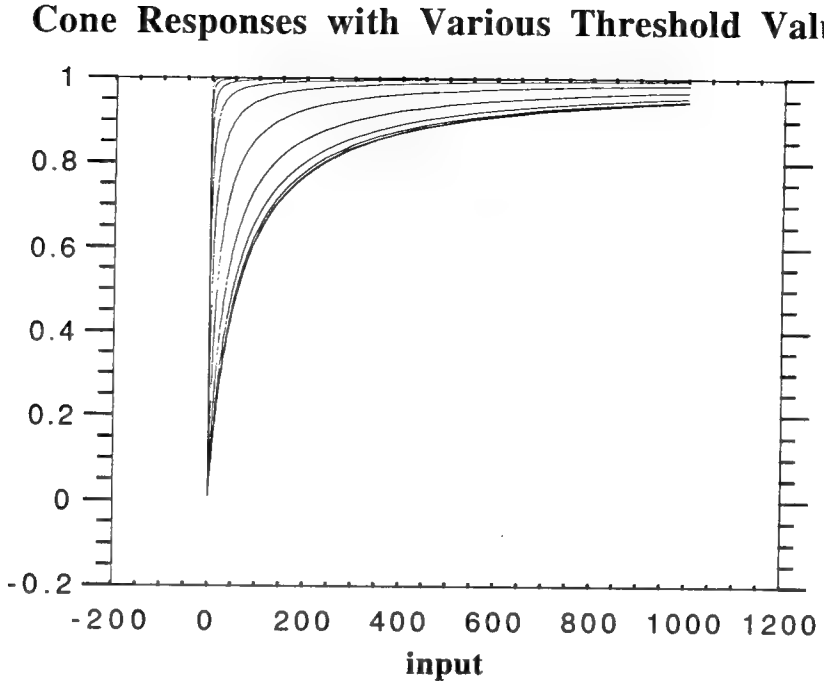
parameter	description	factors in choice of parameter value
ω_{ij}	connection strength between cells	spatial weighting chosen to create receptive field shapes found physiologically, amplitude chosen to keep all stages within linear range of response function
σ_i	threshold of cell i	chosen so that most inputs fall in middle of cell's response range, cone threshold changes with adaptation state
β_i	slope of linear portion of cell's response function	chosen in combination with σ_i to give the appropriate dynamic range for each processing stage
θ	width of adaptation weighting function	small value for simulation of fixation or short presentation time of image, large value for viewing conditions with free eye movements (θ=3.0, relatively small compared to cortical silent surrounds, large compared to center of spectrally opponent receptive fields; $0<\theta<$diameter of image)
α	fraction of total long term adaptation achieved	dependent upon length of viewing time (α=0.2; $0\leq\alpha\leq1$)
c_1, c_2	coefficients for push-pull mechanism	chosen together with a to give a total average constancy shift of 20% or greater for nearly all reflectance-illuminant combinations ($c_1=c_2=0.25$; $0\leq c_1\leq1$; $0\leq c_2\leq1$)


Cortical mechanisms
center input
surround input
output
positive contrast
negative contrast
reference
Interneurons
Spectrally and spatially opponent cells
on-center cells
off-center cells

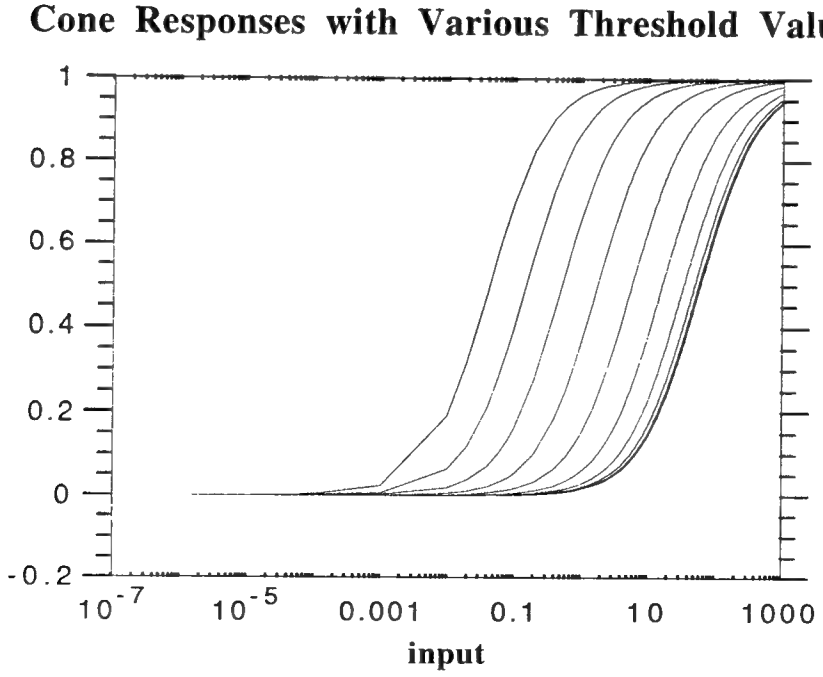

Retinal mechanisms
Spatio-temporal integration
Spectrally and Spatially Opponent Cells
on-center
off-center
Cones
Change of Threshold (Adaptation)

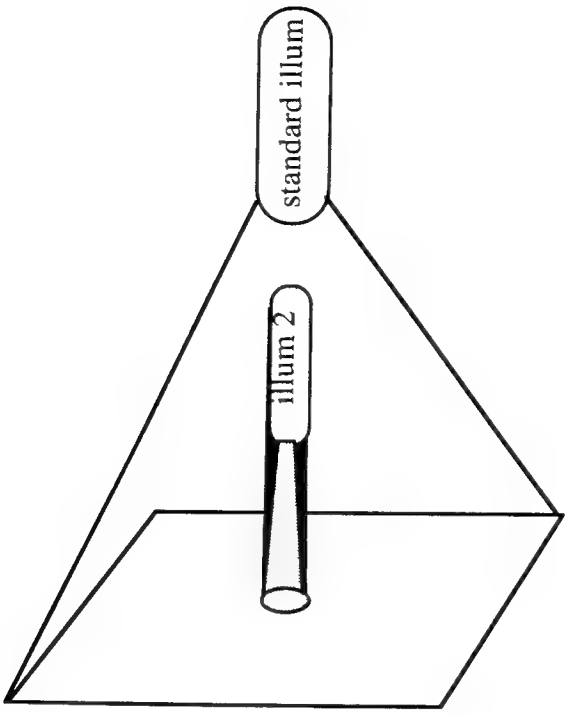

3
2
1
log Relative Sensitivity
R(λ)
B(λ)
G(λ)
400
500
600
700
Wavelength (nm)

(a)

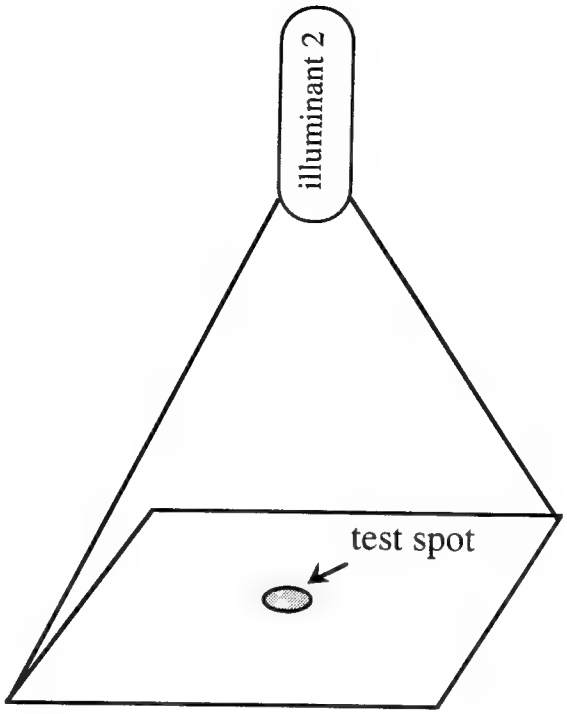

Cone Responses with Various Threshold Values
1
0.8
0.6
0.4
0.2
0
-0.2
-200
0
200
400
600
800
1000
1200
input

(b)


Cone Responses with Various Threshold Values
1
0.8
0.6
0.4
0.2
0
-0.2
10^-7
10^-5
0.001
0.1
10
1000
input

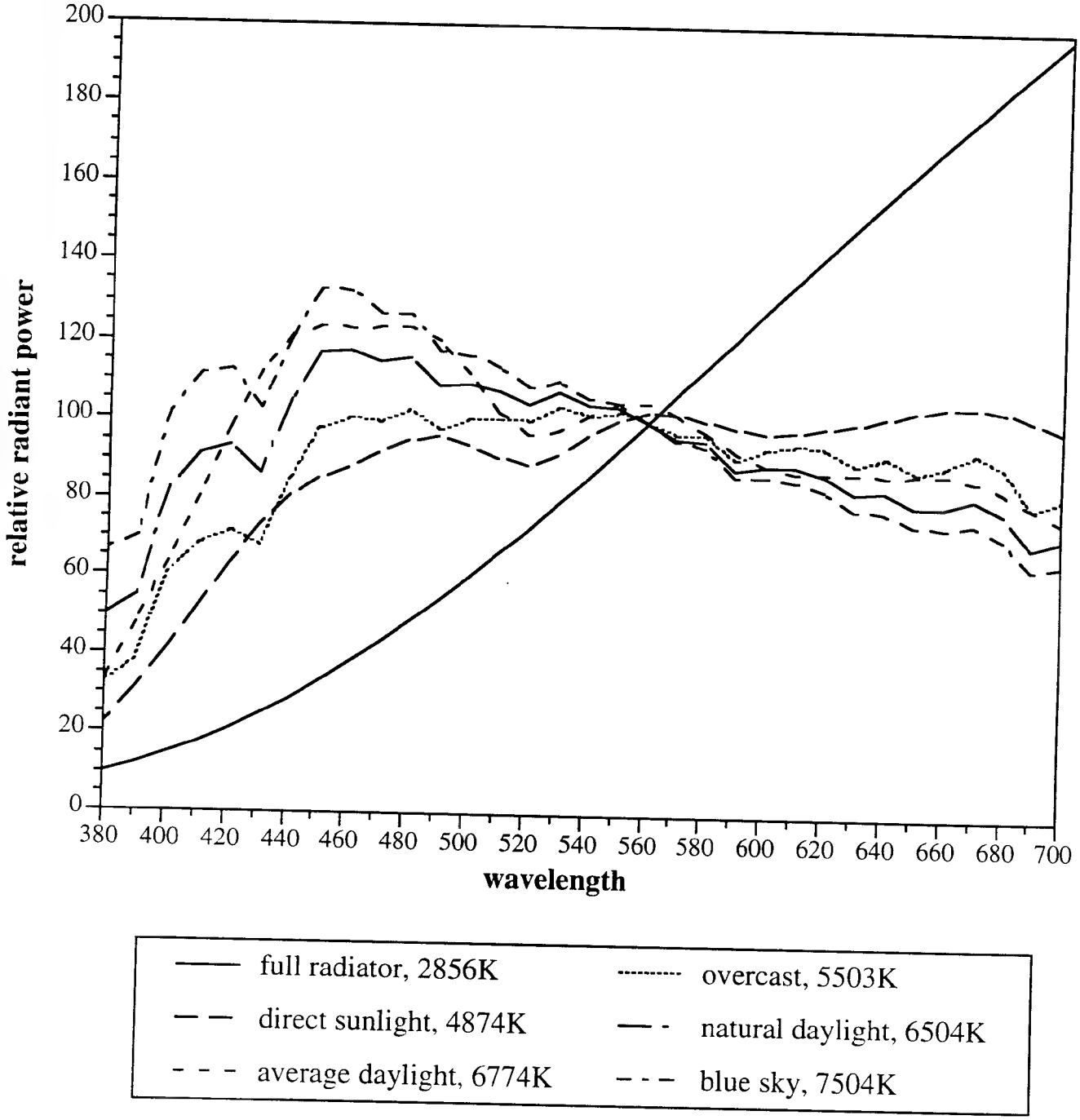


Aperture viewing condition



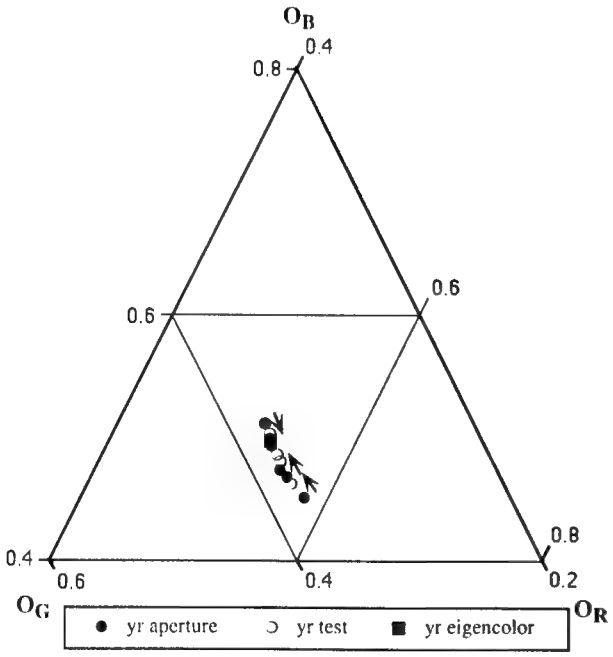
Natural viewing condition,
for test stimuli

Power Spectra of Six Natural Illuminants

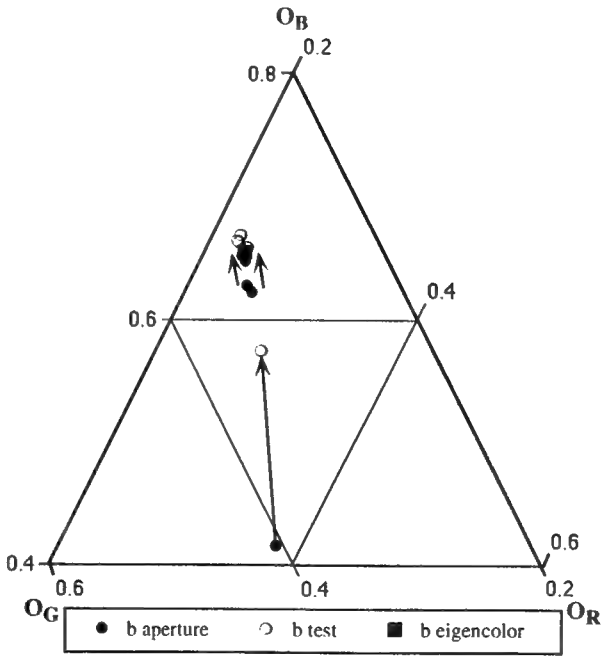


Color Constancy with Natural Illuminant Changes

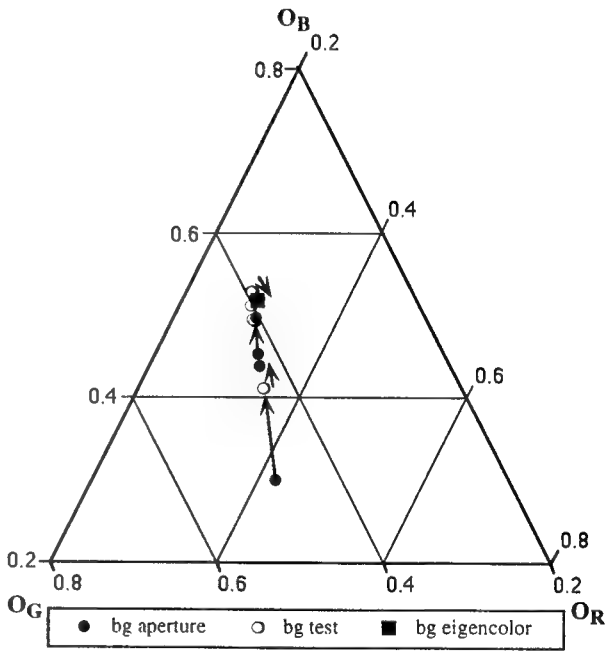
Yellow-Red Reflectance



Blue Reflectance

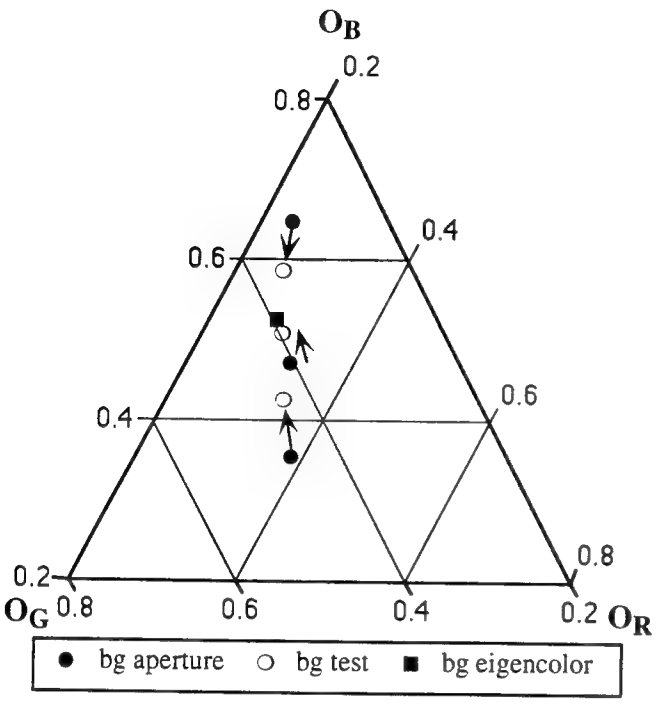


Blue-Green Reflectance

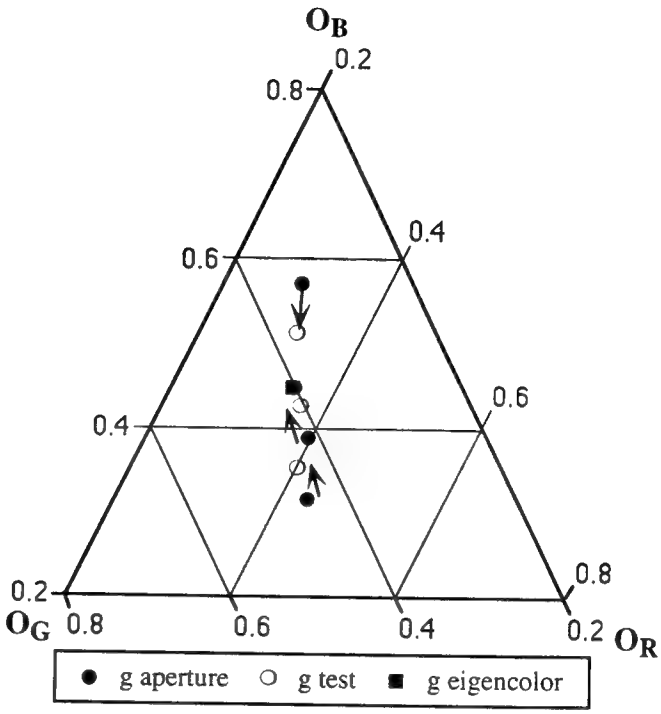


Color Constancy with Larger Illuminant Changes

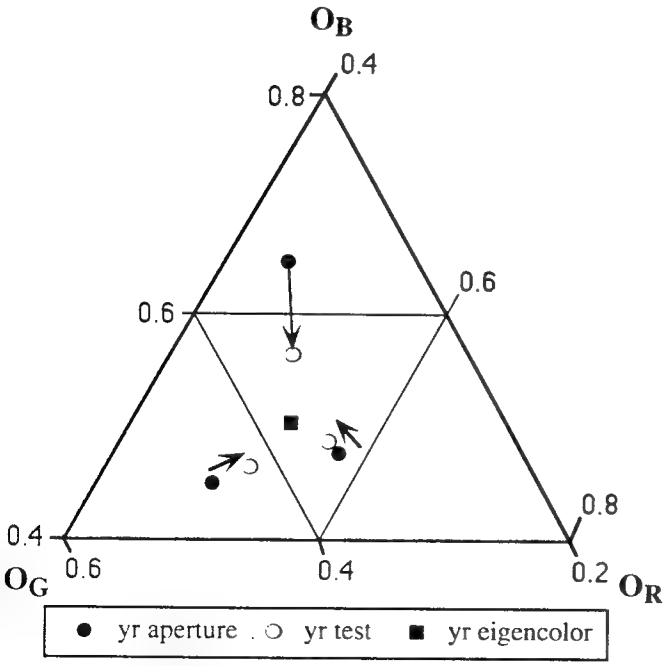
Blue-Green Reflectance



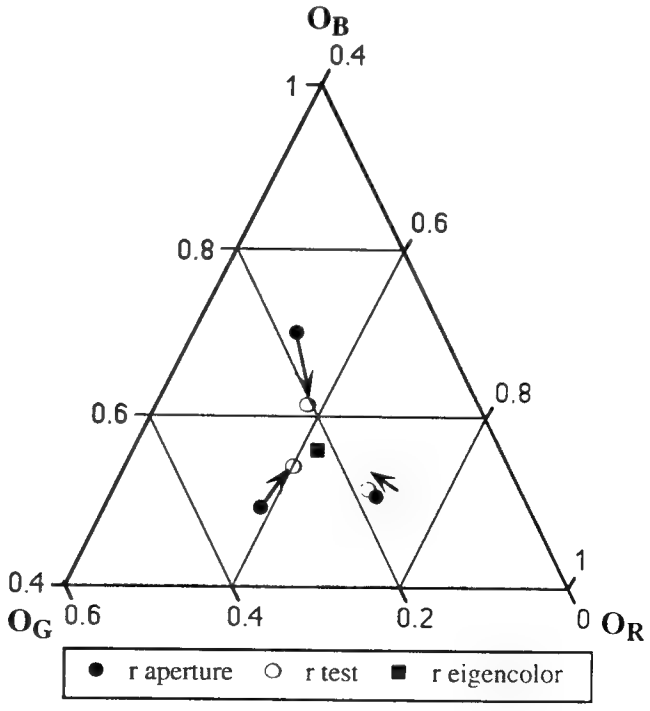
Green Reflectance



Yellow-Red Reflectance

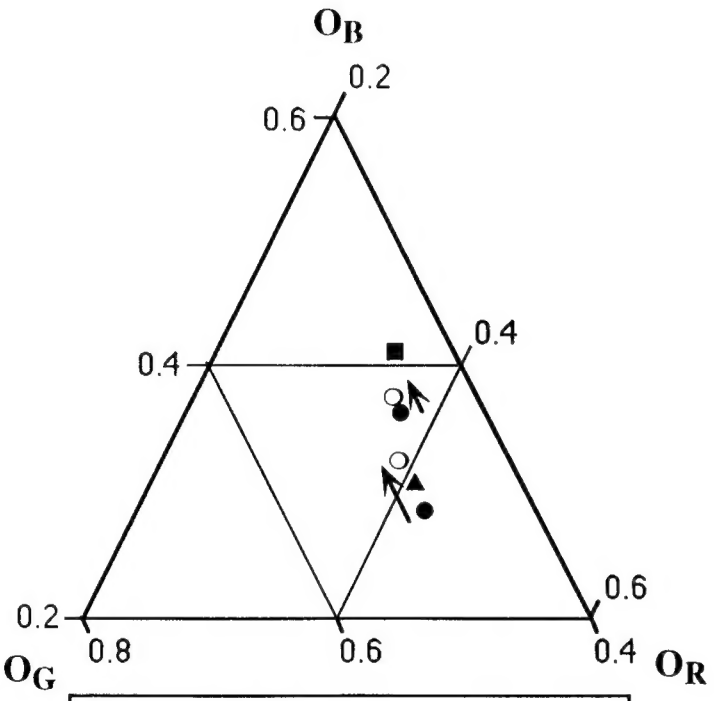


Red Reflectance



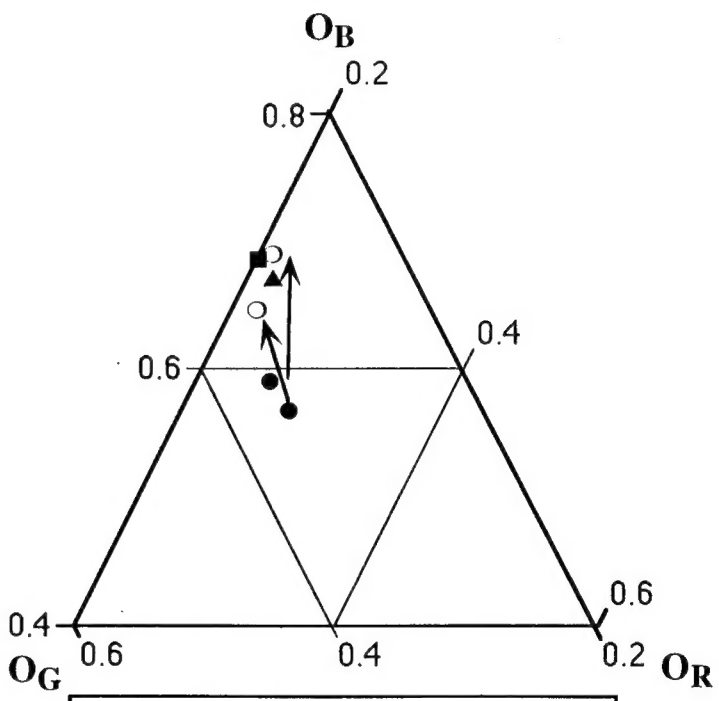
Color Constancy with a Mondrian Background

Green-Yellow Reflectance
and Illuminants A, and B



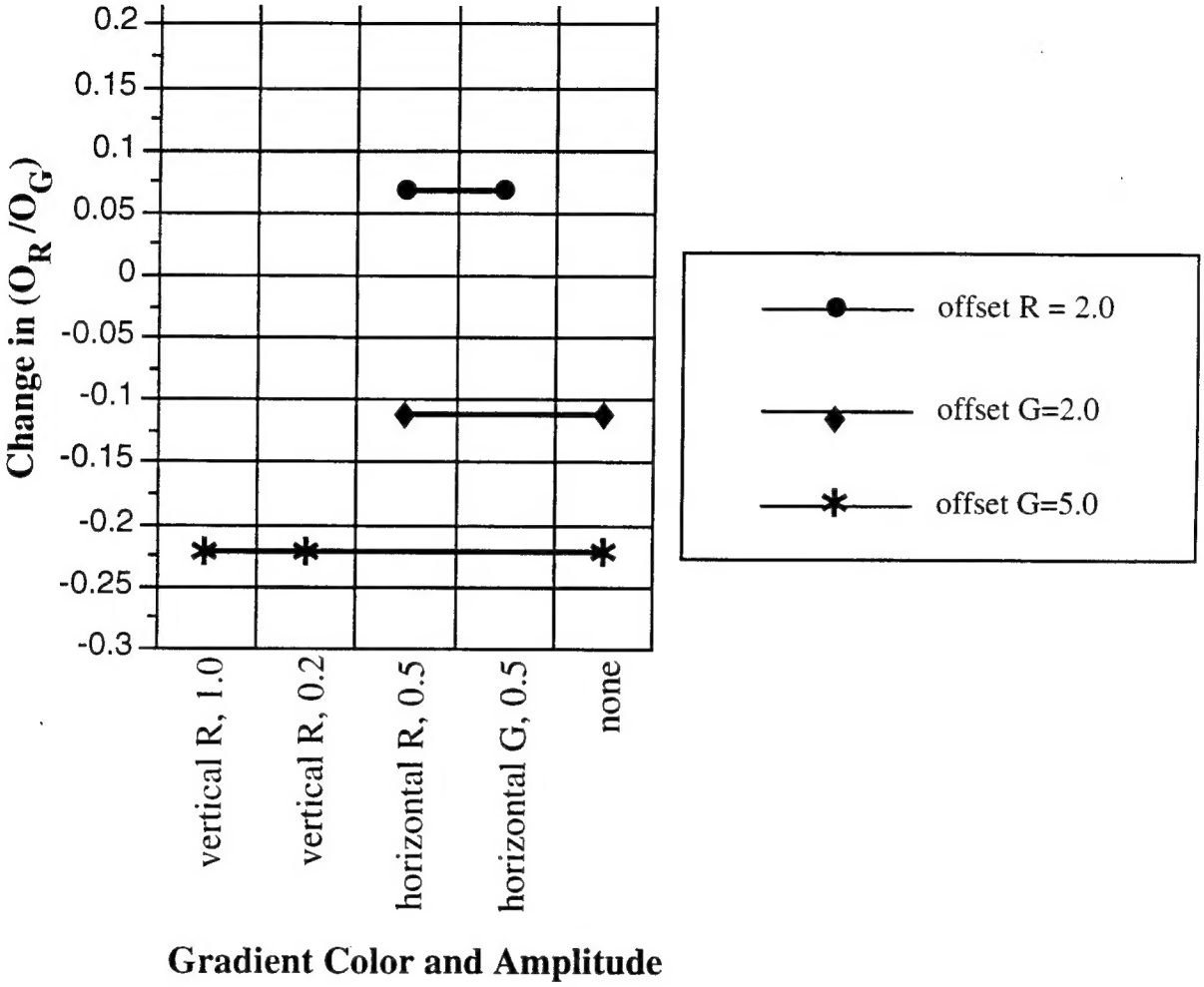
- ● gy aperture
- ○ gy test
- ■ gy neutral illum, mondrian bkgd
- ▲ gy eigencolor

Blue Reflectance
and Illuminants A, and D75

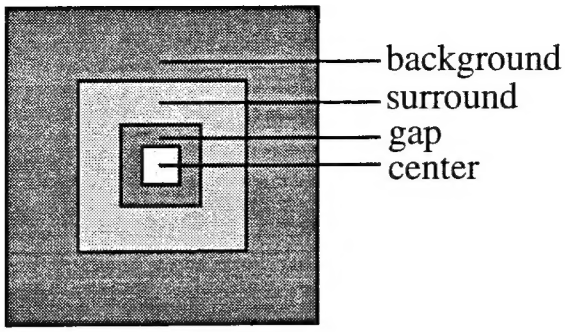


- ● b aperture
- ○ b test
- ■ b neutral illum, mondrian bkgd
- ▲ b eigencolor

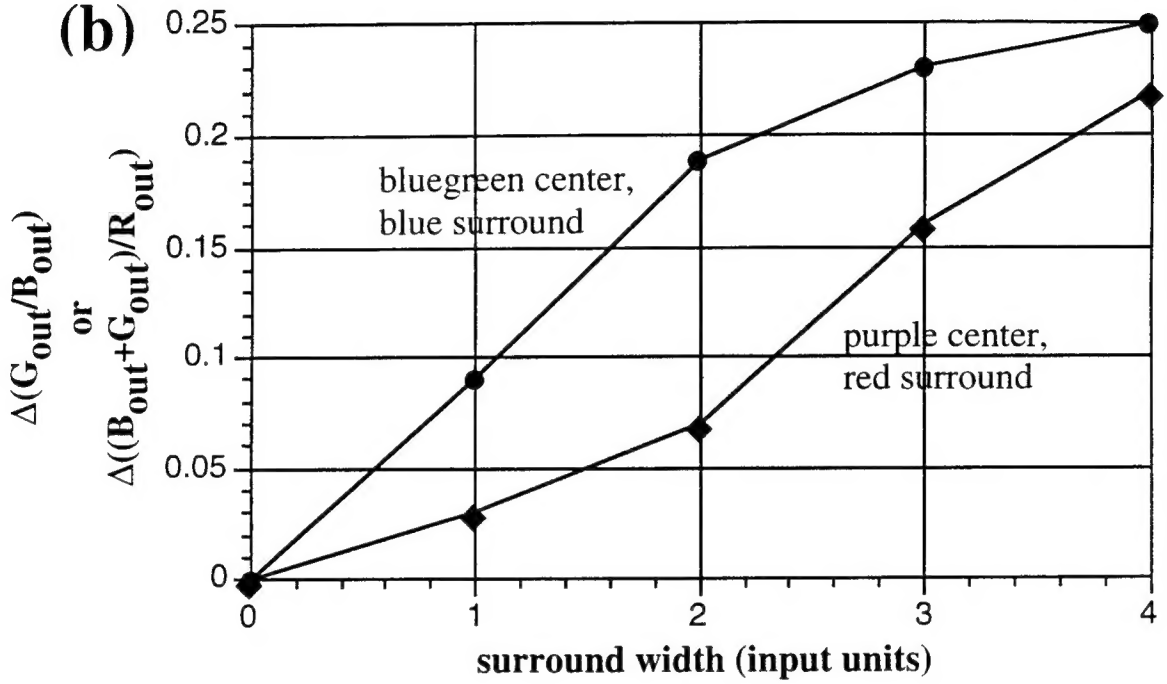
Changes in Output due to
Chromatic Gradients and Offsets in Illuminant



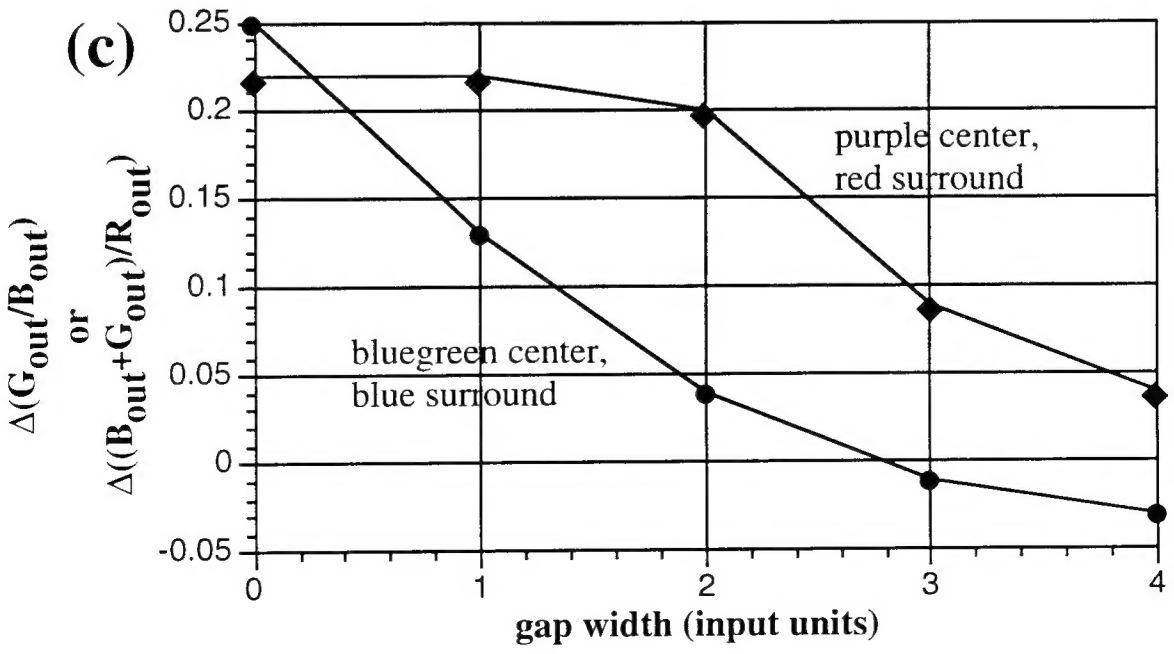
(a)


background
surround
gap
center

(b)

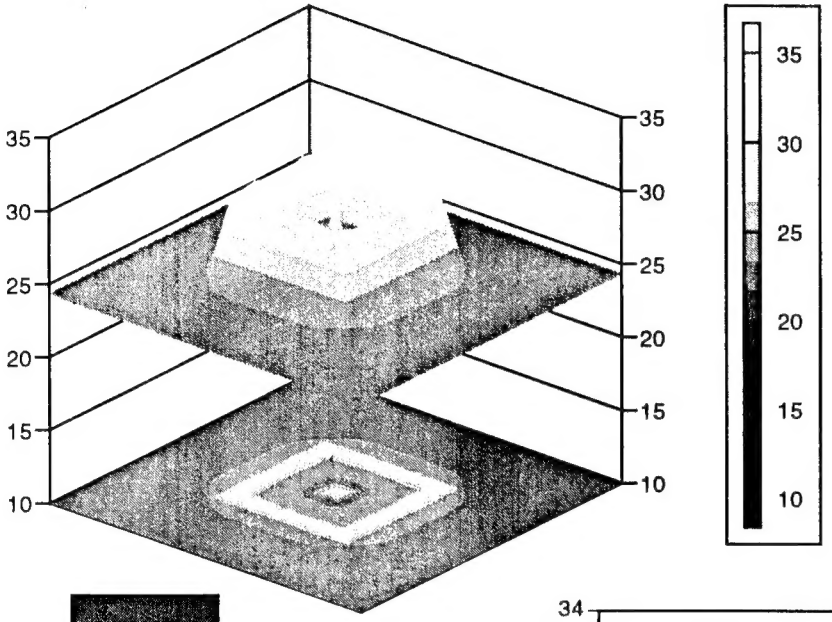

0.25
0.2
0.15
0.1
0.05
0
0
1
2
3
4
bluegreen center,
blue surround
purple center,
red surround
Δ(Gout/Bout)
or
Δ((Bout+Gout)/Rout)
surround width (input units)

(c)

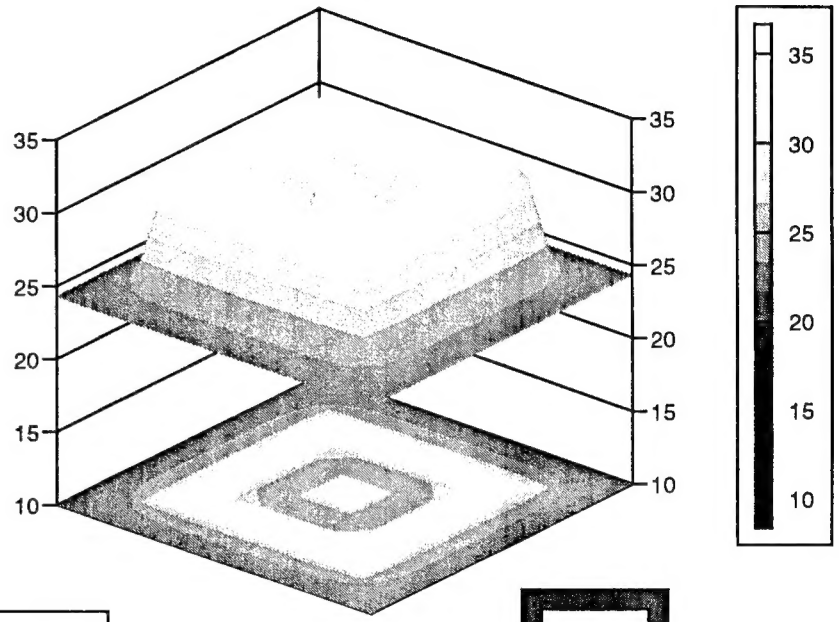

0.25
0.2
0.15
0.1
0.05
0
-0.05
0
1
2
3
4
purple center,
red surround
bluegreen center,
blue surround
Δ(Gout/Bout)
or
Δ((Bout+Gout)/Rout)
gap width (input units)

Assimilation in Fine Color Patterns

On-center R response to Fine Red/Yellow Pattern

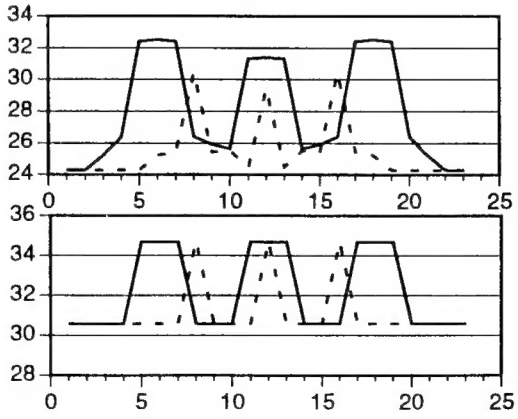


On-center R response to Coarse Red/Yellow Pattern



fine pattern stimulus

response profile



input profile

coarse pattern stimulus